

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Penguin fait peur... Et Google fait tout pour cela ! La stratégie de FUD (fear, uncertainty and doubt) du moteur de recherche fait en sorte que les webmasters craignent chaque lancement du filtre de nettoyage de Google. Mais il faut calmer la parano ambiante et bien comprendre qu'un site web sérieux n'a rien à craindre du palmipède de Mountain View. Voici pourquoi...

Le 22 Mai 2013, la nouvelle version du filtre antispam Google Penguin a été lancée par le moteur de recherche, provoquant une fois de plus des modifications importantes au niveau du positionnement de certains sites sur le moteur de recherche.

Comme à chaque nouvelle mouture, une partie des référenceurs et webmasters ne comprennent pas pourquoi leur site est impacté et craignent les futures mises à jour de Google. Mais en réalité, vous ne devriez pas avoir peur de Google Penguin, et cela pour plusieurs raisons :

- Google est parvenu à faire paniquer les référenceurs et webmasters grâce à sa communication, alors qu'il pénalisait depuis bien longtemps déjà les sites jugés non pertinents ;
- La majorité des sites touchés le sont pour une bonne raison (liens non naturels, contenus pauvres, faible "autorité"...)
- Les sites qui savent se concentrer sur l'utilisateur et leur marque ont profité de cette mise à jour.

C'est quoi, Google Penguin ?

Commençons par la base, à savoir comprendre de quoi nous allons parler : Google Penguin est un filtre appliqué à l'algorithme du moteur de recherche Google.

Celui-ci a été annoncé pour la première fois le 24 Avril 2012. Son but est de décroître la popularité des sites Internet qui ne respectent pas les règles édictées par Google, et qui font donc appel plus ou moins fortement à des techniques interdites ("Black Hat"). Ce filtre cherche donc à diminuer l'efficacité des méthodes suivantes :

- Keyword Stuffing (l'ajout de mots clés partout pour se positionner) ;
- Cloaking (fournir une page différente aux visiteurs et aux moteurs de recherche) ;
- Liens factices et "non naturels" ;
- Contenu caché ;
- Duplication de contenu (notamment le vol de contenu).

La version 2.0 (son nom officiel en interne) de Google Penguin du 22 Mai dernier est en réalité la 4ème version du filtre, dont voici un rapide historique :

- La version 1 sortie le 24 Avril 2012 ;
- La version 1.1, une mise à jour mineure le 26 mai 2012 (appelé Penguin 2 par les référenceurs) ;
- La version 1.2, une seconde mise à jour mineure le 5 octobre 2012 (appelée Penguin 3) ;
- La version 2 fut une mise à jour importante le 22 Mai 2013 (appelée Penguin 4).

Il s'agit bel et bien d'un filtre et non pas d'un algorithme, et la différence entre les deux est cruciale : un filtre va s'appliquer à un instant T, en fonction des données que le moteur de recherche aura à sa disposition. A l'inverse, un algorithme se calcule constamment, jour après jour. Et c'est justement parce qu'il s'agit d'un filtre que l'on ressent plus fortement ce type de mise à jour, car son impact est immédiat et touche un grand nombre de sites internet différents à la fois (sources

<http://www.abondance.com/actualites/20130523-12635-google-penguin-4-ete-lance.html>
et <http://www.faire-un-site.com/informations-sur-pengin.php>).

C'est d'ailleurs le contraire avec l'autre filtre très connu Google Panda, qui fait maintenant partie intégrante de l'algorithme de Google

"Matt Cutts, Google's head of search spam, announced at SMX West today that their Panda algorithm will soon be more integrated into their overall continuous algorithm updates."

Source : <http://searchengineland.com/google-panda-to-be-integrated-into-the-search-algorithm-panda-everflux-151528>

Penguin 2.0

La V1.1 et 1.2 étaient ce qu'on appelle de simples "Data Refresh", c'est à dire une mise à jour des données pour le filtre concerné. La V2 est différente dans le sens où c'est justement l'algorithme qui est modifié, tout en mettant également à jour les données utilisées. Penguin 2.0 devrait donc avoir plus d'impact et de visibilité auprès des webmasters que les deux précédentes mises à jour.

Selon une vidéo récente, voici ce qu'annonçait Matt Cutts par rapport au filtre 2.0 de Penguin et des nouveautés que son équipe allait développer dans les prochains mois :

- une pénalisation pour l'utilisation trop importante des publireportages et communiqués de presse (popularité non naturelle) ;
- réduire l'efficacité des liens considérés comme du "spam" ;
- détecter plus efficacement les sites ayant des malwares et virus ;
- réévaluer à la hausse le classement des sites d'autorité.

Source : <http://www.mattcutts.com/blog/what-to-expect-in-seo-in-the-coming-months/>

Qui est pénalisé ?

Google ne fournissant pas de liste lui-même, plusieurs agences ont publié leurs statistiques sur les sites pénalisés par Google Penguin 2.0, dont voici un exemple :

Domain	SEO Visibilitybefore	SEO Visibility after	loss	in %
metrofrance.com	93 581	8 672	-84 909	-90,73
filmdoz.com	9 939	2 100	-7 839	-78,87
v2q.com	31 433	7 473	-23 960	-76,23
annuaire.pro	34 863	9 051	-25 812	-74,04
jeuxplus.com	10 238	2 876	-7 362	-71,91
bonnevideox.info	16 872	4 825	-12 047	-71,40
bonporn.com	73 345	21 012	-52 333	-71,35
alloemploi.fr	13 087	4 202	-8 885	-67,89
square-annuaire.com	21 004	6 921	-14 083	-67,05
centralepneus.fr	12 987	4 442	-8 545	-65,80
wondershare.fr	10 412	3 614	-6 798	-65,29
ooyyo.fr	11 689	4 087	-7 602	-65,04
videocomplete.info	10 030	3 537	-6 493	-64,74
sexeteub.com	23 687	9 038	-14 649	-61,84
jeux-1.com	15 302	5 943	-9 359	-61,16
sitacados.com	15 497	6 161	-9 336	-60,24
frmovies.net	12 338	5 127	-7 211	-58,45
midilibre.com	17 242	7 423	-9 819	-56,95
jekt.fr	23 670	10 692	-12 978	-54,83
hardsextube.com	14 269	6 517	-7 752	-54,33
sexezamateur.com	16 422	7 888	-8 534	-51,97
pricededuc.com	18 463	8 923	-9 540	-51,67
jeuxvideo-flash.com	171 457	85 357	-86 100	-50,22
jeuxdroles.fr	25 395	12 696	-12 699	-50,01
pornego.com	15 722	7 900	-7 822	-49,75
streamingclic.com	14 181	7 157	-7 024	-49,53
letelegramme.fr	66 431	33 630	-32 801	-49,38
videobuzzy.com	20 843	10 738	-10 105	-48,48
divertissez-vous.com	26 526	14 072	-12 454	-46,95
grandlyon.com	13 898	7 384	-6 514	-46,87
1001actus.com	21 566	11 500	-10 066	-46,68
sexetag.com	50 104	26 781	-23 323	-46,55
kporno.com	30 327	16 344	-13 983	-46,11
persun.fr	18 851	10 226	-8 625	-45,75
extrait-video-x.com	16 275	9 133	-7 142	-43,88
jetcost.com	27 034	15 323	-11 711	-43,32
jeux.fm	55 749	32 559	-23 190	-41,60
vkstreaming.com	16 438	9 614	-6 824	-41,51
agentdunet.com	35 669	20 922	-14 747	-41,34
p0rno.com	19 492	11 485	-8 007	-41,08

Source : <http://www.abondance.com/actualites/20130530-12677-la-liste-des-sites-punis-par-penguin-4-en-france.html>

Attention : pour les sites *metrofrance* et *midilibre*, leur présence dans le tableau s'explique par le fait que les noms de domaines ont été modifiés depuis.

Nous ne pouvons malheureusement tirer aucune conclusion de tels chiffres, mais nous pouvons affirmer la chose suivante : un site impacté par le filtre Google Penguin est souvent fortement pénalisé d'un seul coup, d'où une partie de la peur que peuvent avoir certains webmasters et référenceurs à ce sujet.

SeoMoz (renommé Moz depuis peu) a, lui aussi, publié ses propres statistiques. La partie la plus intéressante de leur étude concerne les 20 "ténors" du web qu'ils ont analysé avant et après Google Penguin 2.0. Et le résultat est relativement logique : très peu de variations, souvent comprises entre +1% ou -1%. En d'autres termes, les sites qui se positionnaient très bien sur de nombreuses requêtes, et qui sont donc reconnus par Google comme des sites de qualité (Wikipédia, Ebay, Amazon, ...), ne sont absolument pas impactés par ce type de filtre.

Impact de Google Penguin 2.0 sur 20 ténors du web	
Domaines	"Part de marché SERP" et impact de Penguin 2.0
Source : SeoMoz.org	
1. en.wikipedia.org	5.66% (+0.29%)
2. www.amazon.com	2.35% (-0.75%)
3. www.youtube.com	2.22% (+3.11%)
4. www.facebook.com	1.49% (+6.05%)
5. www.yelp.com	1.35% (-8.11%)
6. twitter.com	0.84% (+4.77%)
7. www.webmd.com	0.58% (+0.37%)
8. pinterest.com	0.58% (+1.87%)
9. www.walmart.com	0.52% (+1.24%)
10. www.tripadvisor.com	0.49% (+4.54%)
11. www.foodnetwork.com	0.47% (+0.45%)
12. allrecipes.com	0.47% (-0.44%)
13. www.ebay.com	0.44% (+1.98%)
14. www.mayoclinic.com	0.41% (-0.76%)
15. www.target.com	0.38% (+1.72%)
16. www.yellowpages.com	0.37% (-4.37%)
17. popular.ebay.com	0.37% (+0.58%)
18. www.huffingtonpost.com	0.36% (+2.12%)
19. www.overstock.com	0.33% (+3.27%)
20. www.indeed.com	0.32% (-0.32%)

Source : <http://www.journaldunet.com/solutions/seo-referencement/l-impact-de-penguin-2-0-en-chiffres.shtml>

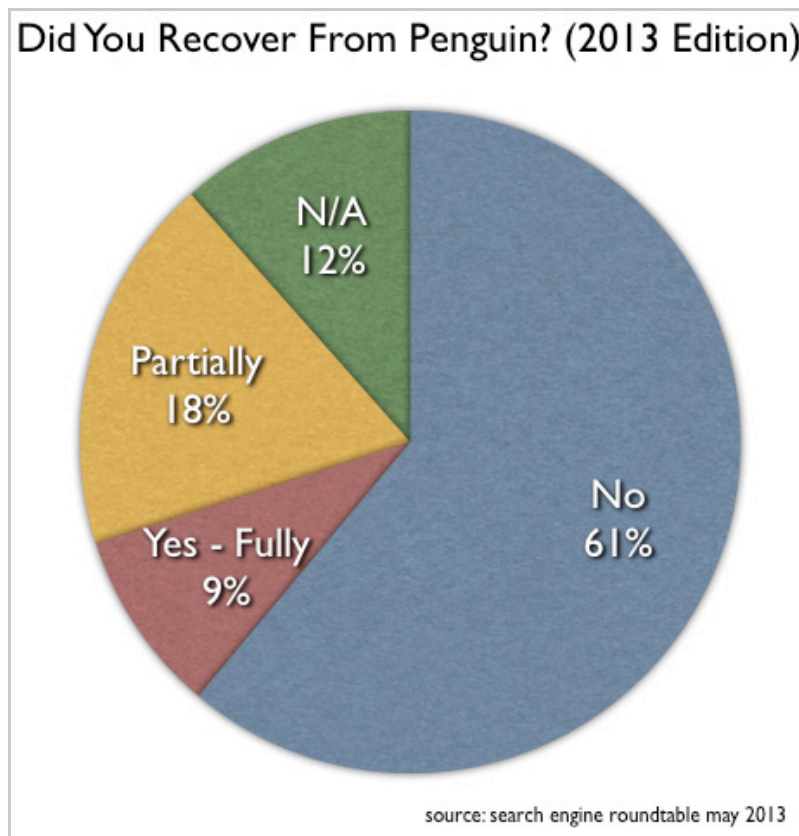
Webrankinfo et d'autres sites semblent confirmer la tendance : plus un site aura travaillé sa marque, sa présence sociale et ses auteurs, plus il profitera de la mise à jour de Google Penguin, tandis que les sites faisant l'inverse auront tendance à être pénalisés.

Source <http://www.webrankinfo.com/dossiers/etudes/perdants-pingouin-4>

Comment fonctionne Penguin ?

Le gros problème de Google Penguin est surtout qu'il s'agit d'un filtre appliqué à un instant T, plutôt qu'un changement constant de la manière dont Google classe les sites Internet. Une petite étude du site Serountable datée du 15 mai 2013 montre qu'il est

possible de récupérer ses positionnements après avoir été touché par Google Penguin, mais que peu de sites y parviennent (les données sont basées sur un sondage ayant recueilli 500 réponses) :



Source : <http://www.seroundtable.com/google-penguin-poll-16788.html>

D'ailleurs, ce résultat est assez logique car il faut souvent attendre une réévaluation de la part des équipes de Google ou la prochaine mise à jour de Google Penguin pour éventuellement sortir de la pénalité, comme l'ont remarqué plusieurs référenceurs, notamment avec cette version 2 de Google Penguin :

- Marie Pourreyron : "Ce site était très bien positionné sur plusieurs requêtes rémunératrices, mais, l'année dernière, Penguin l'a fait plonger dans les abîmes des résultats de Google. Je l'ai totalement abandonné, mais il est revenu hier en deuxième page sur deux mots clés" ;
- "Sylvain Richard a aussi vécu exactement la même expérience, avec un site sévèrement, mais pertinemment pénalisé par Penguin l'année dernière, puis abandonné. Ce Penguin 2.0 l'a aussi fait réapparaître dans la deuxième page."

Source : <http://www.journaldunet.com/solutions/seo-referencement/seo-impact-de-google-penguin-2-0.shtml>

Pourquoi ne faut-il pas avoir peur ?

Sur le papier, ce filtre fait peur : les équipes de Google communiquent fortement sur ce sujet, conjointement avec de très nombreux sites qui relaient l'information dès qu'une mise à jour est en ligne. Et pourtant, il y a de nombreuses raisons qui doivent vous pousser à ne plus craindre Google Penguin.

La communication

La première, c'est que le tapage médiatique réalisé autour de ce changement d'algorithme est entièrement orchestré par les équipes de Google. Bien avant Penguin ou

Panda, Google pénalisait déjà les sites qui utilisaient des techniques non conformes (quand il les détecte bien entendu).

La principale différence est que désormais Google fait le ménage à date fixe, et il pénalise ainsi de très nombreux sites Internet le même jour partout dans le monde, contrairement à avant où cela s'étalait dans la durée et où la firme ne communiquait pas autant sur ce sujet.

En d'autres termes, Google est parvenu à instaurer une psychose dans le monde du référencement, justement pour pousser les webmasters à travailler au maximum leurs contenus et à éviter d'utiliser toute technique de référencement jugée "abusive" ou "non naturelle". Rien n'a donc changé : Google cherche et cherchera toujours à proposer le meilleur contenu pour chaque requête, en essayant de combattre toute technique non naturelle de référencement.

Un nombre de requêtes faible

L'autre raison, c'est qu'en réalité le nombre de requête impacté est relativement faible. Pour cette version 2 par exemple, seul 2,3% des recherches en anglais sont affectées (source : <http://www.matcutts.com/blog/penguin-2-0-rolled-out-today/>).

En résumé, cela touche peu de mots clés et expressions au final, même si ceux qui sont touchés le sont par contre de manière importante et drastique.

Quand on est pénalisé, c'est pour une bonne raison !

Troisième point, et sans doute le plus crucial : la plupart des sites touchés le sont malheureusement pour une bonne raison. Quand un site Internet est touché par la mise à jour de Google Penguin, une analyse permet souvent de mettre en avant des points communs qui sont tout sauf naturels :

- des liens ayant trop souvent une ancre exacte ;
- des liens presque intégralement en dofollow ;
- un contenu dupliqué, pauvre et peu pertinent ("bouillie" de contenus, vol de publications, ...) ;
- des sites récents ayant une faible autorité dans leur domaine, ainsi que sur Internet en général ;
- des liens en provenance de sites ayant ce même profil.

Nous pourrions résumer cela en quelques mots : du spam, du contenu inutile et des liens artificiels.

La différence entre réalité et pénalité

Attention, nous avons cependant tous remarqué que certains sites utilisant ces techniques n'ont pas été touchés. Et cela peut s'expliquer au travers d'une question simple : Google a-t-il eu le temps et/ou la possibilité de détecter les techniques utilisées par chaque site dans son index ?

Puisque Google doit parcourir et indexer tout le web mondial, il est logique qu'il y ait un décalage entre l'application de ses filtres et algorithmes et la réalité des données en ligne sur chaque site Internet.

C'est d'ailleurs également vrai avec les liens que vous allez créer : leur impact ne sera que rarement immédiat, car il faudra laisser plus ou moins de temps à Google pour qu'il les prenne en compte.

Travaillez votre référencement

Pour ne plus craindre ce type de mise à jour ou une éventuelle pénalité manuelle, il est fortement recommandé de revenir à la base du métier de référenceur : l'internaute.

Les sites non impactés sont souvent des sites de qualité, avec un vrai contenu unique qui répond au besoin de l'internaute. Ils sont souvent reconnus par les autres sites du même secteur d'activité, sont présents sur les réseaux sociaux et publient régulièrement du contenu. En d'autres termes, ils sont centrés sur l'utilisateur et leur marque, et non pas sur le moteur de recherche.

Sur quoi faut-il se concentrer ?

La clé est donc de revenir à la base du référencement naturel, à savoir :

- un contenu unique et pertinent ;
- un contenu publié ou mis à jour régulièrement ;
- un site bien conçu et bien structuré ;
- des liens créés de manière naturelle (le plus possible, en tout cas) ;
- des gens et une société réelle derrière le site.

Vous devez donc gérer votre site comme une entité réelle, à savoir avec un renouvellement constant de votre activité. La fraîcheur des contenus étant importante chez Google, il faut montrer que son site est actif.

Et de ce fait, il faut également prouver qu'il y a des personnes réelles derrière le site Internet, et qui veulent apporter des réponses aux besoins et questions des internautes. C'est là qu'entre en jeu l'authorship et les réseaux sociaux.

Bien entendu, la base de tout reste et restera vos contenus et la structure de votre site. Soyez pertinents et répondez aux besoins de l'internautes, et votre site devrait s'en sortir sans aucun problème !

Ce qu'il faut éviter

D'ailleurs, les sites qui ont été pénalisés par Google font souvent appel aux techniques suivantes qui sont à l'opposé de ce qui vient d'être dit (source :

<http://support.google.com/webmasters/bin/answer.py?hl=fr&answer=35769#3>) :

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

Google avait d'ailleurs lancé une page officielle pour expliquer clairement ce qui pose problème selon lui. Le lien officiel est celui-ci :

<https://sites.google.com/site/webmasterhelpforum/fr/fr-posts/une-autre-mesure-pour-recompenser-les-sites-de-haute-qualite>

Conclusion

Google est parvenu à créer une psychose dans le monde du référencement naturel, dans le but de réduire toutes les techniques qu'il juge contraire à ses règles. Mais Google pénalise et traque les liens et contenus artificiels depuis des années, c'est uniquement la méthode de déploiement des pénalités qui évolue et qui fait justement parler d'elle.

Google Penguin fait donc peur, car son impact est immédiat et surtout généralisé sur l'ensemble du web mondial.

Et pourtant, la plupart des sites pénalisés le sont pour une raison simple : ils ont oublié l'essence même du référencement naturel : l'internaute et les contenus qui répondent à ses besoins. Des valeurs qu'il faudra bien un jour remettre au centre de votre visibilité web !

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