

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Cet article analyse, grâce à de nombreux chiffres, comparatifs et exemples inédits, l'impact qu'a eu la mise à jour Panda de l'algorithme de Google en Grande-Bretagne. Que s'est-il réellement passé ? Quels types de site ont été les plus touchés : comparateurs de prix, fermes de contenus, annuaires, sites d'avis et de questions/réponses, sites d'annonces, agrégateurs de contenus, de coupons ou moteurs de recherche ? Et quels sites en ont bénéficié ? Et si un impact fort a été détecté, en négatif ou en positif, pourquoi ? Cette analyse permet de tirer certaines conclusions importantes et intéressantes, surtout au moment où le déploiement de Panda est attendu en France très prochainement ...

Google a déployé, il y a quelques semaines de cela, une mise à jour majeure de son algorithme, baptisée Panda, dans tous les index anglophones de son moteur de recherche. L'impact de cette modification sur les classements a été sans précédent : selon les propres chiffres communiqués par Google, Panda a modifié 12% des requêtes sur la version US du site ! Il faut probablement remonter à la mise à jour Florida, le 16 novembre 2003, pour retrouver un changement aussi radical dans les page de résultats de Google.

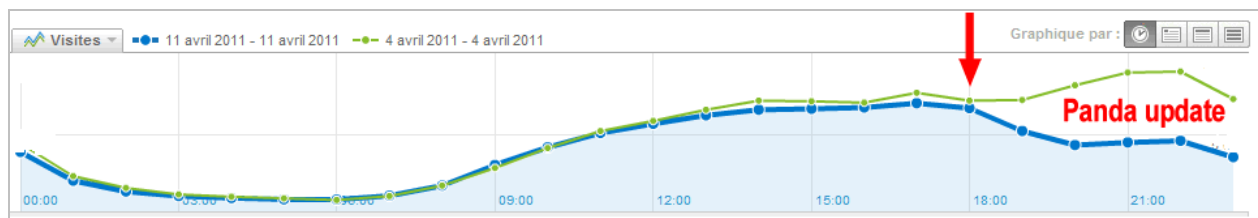
Un article très complet de Guillaume Thavaud décrivant cette mise à jour et les premiers constats effectués par les communautés de référenceurs anglo-saxons a déjà été publié dans le numéro d'avril 2011 de la lettre "Recherche & Référencement". Nous invitons nos lecteurs à s'y référer pour prendre connaissance de l'historique de la conception et du déploiement de Panda, points sur lesquels nous ne reviendrons pas ici.

Nous disposons maintenant d'un recul de plus de deux mois sur la mise à jour Panda aux USA, et de presque un mois pour les autres index anglophones. Il nous a paru intéressant, à quelques jours (ou semaines ?) du déploiement dans l'espace francophone, d'analyser plus en détail la nature et l'impact de cette modification majeure de l'algorithme. D'autant plus que la deuxième phase du déploiement, le 11 avril 2011, a permis de déceler des caractéristiques nouvelles de l'algorithme.

Grâce à des données exclusives fournies par l'agence de référencement Aposition, et grâce à l'accès qui nous a été fourni par plusieurs webmasters de sites anglo-saxons, nous allons pouvoir jeter un regard nouveau sur cette mise à jour, et savoir s'il faut vraiment en craindre ou en espérer les conséquences sur les sites francophones.

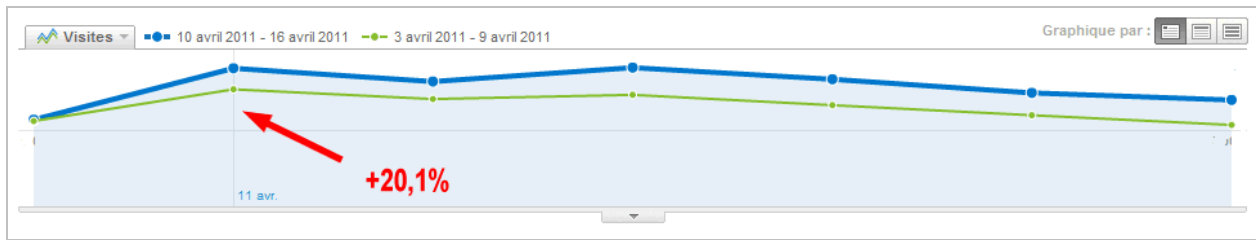
L'impact réel de Panda dans sa version du 11 avril

Le graphique ci-dessous montre l'effet immédiat de la mise à jour sur le trafic d'un site britannique : la journée semble s'annoncer normale, lorsque soudain, à 18h00 BST, le trafic décroche violemment !



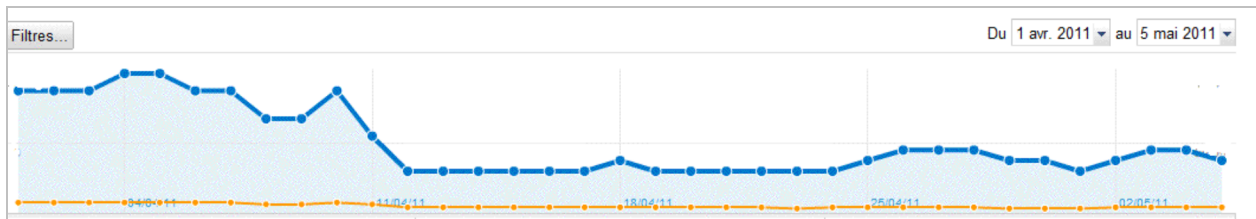
Comparaison de l'évolution heure par heure du nombre de visites sur un site britannique, entre le 4 avril 2011 (en vert) et le 11 avril 2011 (en bleu).

Au même moment, sur un autre site (irlandais cette fois), le trafic connaît du jour au lendemain un bond de ... 20% !



Evolution du trafic hors marque, issu du moteur Google, sur un site Irlandais, non touché par Panda.

Si l'on observe les changements de position, on comprend mieux ce qui s'est passé : d'un côté, le premier a reculé de deux ou trois pages dans les résultats de Google sur un grand nombre de requêtes, et même disparu complètement des premières pages. De l'autre, le site a gagné à quelques places, parfois plus de 10, en raison de l'élimination d'un grand nombre de sites en tête des résultats. Sur le deuxième site, qui a bénéficié de Panda, l'impact est net mais limité à certains groupes de requêtes : Panda a plus ou moins d'effet selon les univers.



Ci-dessus, graphe de l'évolution des impressions issu du compte Google Webmaster Tools d'un autre site britannique impacté par Panda.

Quel est le Panda à l'origine du nom de l'algorithme ?

La mise à jour Panda a d'abord été appelée "Farmer" par la communauté SEO, reprenant le nom choisi par Dany Sullivan du site Search Engine Land. "Farmer", parce qu'elle était supposée s'attaquer aux fermes de contenu. On verra dans la suite de cet article que le périmètre des sites ciblés par Panda est en fait bien plus large et diversifié. Google a révélé assez vite que le nom de code interne donné à cette "update" était en fait Panda, et ce nom officiel a été aussitôt adopté par l'ensemble des SEO. Mais pourquoi avoir choisi le nom d'un placide plantigrade amateur de bambous pour cette mise à jour dont le totem aurait pu être aussi bien un grizzli des montagnes, franchement omnivore ?

En fait, les ours n'ont rien à voir avec ce nom de baptême. En réalité, Panda est le nom de baptême d'un ingénieur de Google, à qui on attribue la paternité de la méthode d'évaluation qui a rendu cette mise à jour possible. Cette "révélation" a été faite dans une interview d'Amit Singhal, le responsable de l'algorithme de Google, et de Matt Cutts dans le magazine Wired. Le problème, c'est qu'il existe deux Panda parmi les ingénieurs de Google : Biswanath Panda et Navneet Panda, qui ont un cursus et des sujets d'intérêt très similaires (ce sont tous les deux des spécialistes des programmes d'apprentissage automatique). Certains se sont précipités pour lire les publications scientifiques des deux Panda, afin de voir si elles ne révélaient pas l'approche adoptée pour cette mise à jour mais en vain : leurs publications sont peu nombreuses et leur sujet trop éloigné de l'objectif de l'update Panda pour en tirer des informations utiles. Bien tenté quand même...

Quels sites ont été impactés par la mise à jour Panda ?

Lors du déploiement de la première version de Panda aux USA, la typologie des sites impactés n'était pas claire, et on pouvait encore croire à ce qui se disait dans la communauté SEO US, c'est-à-dire que l'update Panda ciblait essentiellement les fermes de contenu.

Dans le cas des sites anglais, on s'aperçoit que la liste des principaux sites impactés est beaucoup plus diverse et en même temps très reconnaissable.

Le tableau suivant, fourni par l'agence Aposition-Isobar, donne les 40 sites les plus touchés par l'*update* sur un échantillon de 250 000 requêtes. Le classement est effectué en fonction du nombre de positions en première page perdues après la mise à jour.

	Nb expr.				Nb expr. Page 1			
www.reviewcentre.com	12078	24152	12074	49,99%	699	2854	2155	75,50%
news.bbc.co.uk	42241	45241	3000	6,63%	7315	8859	1544	17,42%
www.ciao.co.uk	13082	22277	9195	41,27%	167	1673	1506	90,01%
pt.wikipedia.org	1575	1893	318	16,79%	538	1861	1323	71,09%
fr.wikipedia.org	676	1600	924	57,75%	232	1385	1153	83,24%
www.cylex-uk.co.uk	16662	33700	17038	50,55%	155	1278	1123	87,87%
www.qype.co.uk	16675	27555	10880	39,48%	250	1333	1083	81,24%
www.dooyoo.co.uk	12142	24900	12758	51,23%	81	930	849	91,29%
it.wikipedia.org	777	1085	308	28,38%	248	1010	762	75,44%
es.wikipedia.org	785	1035	250	24,15%	222	963	741	76,94%
www.myvouchercode.co.uk	9398	12579	3181	25,28%	840	1561	721	46,18%
simple.wikipedia.org	1177	1517	340	22,41%	627	1330	703	52,85%
www.kelkoo.co.uk	1301	8446	7145	84,59%	63	756	693	91,66%
www.answers.com	4906	19815	14909	75,24%	277	946	669	70,71%
www.wisegeek.com	2689	8889	6200	69,74%	535	1096	561	51,18%
www.aboutbritain.com	3375	7510	4135	55,05%	100	613	513	83,68%
www.ukpropertyshop.co.uk	3061	3577	516	14,42%	277	765	488	63,79%
www.hotfrog.co.uk	11104	26584	15480	58,23%	60	533	473	88,74%
www.information-britain.co.uk	6871	11308	4437	39,23%	165	619	454	73,34%
www.pixmania.co.uk	932	5700	4768	83,64%	44	479	435	90,81%
www.worldtravelguide.net	1149	5273	4124	78,20%	112	514	402	78,21%
sv.wikipedia.org	352	540	188	34,81%	121	509	388	76,22%
www.comparestoreprices.co.uk	6344	7625	1281	16,80%	58	441	383	86,84%
pl.wikipedia.org	454	554	100	18,05%	178	531	353	66,47%
www.hotelclub.com	2498	5522	3024	54,76%	53	392	339	86,47%
www.buyyourcar.co.uk	3194	4238	1044	24,63%	413	732	319	43,57%
www.londontown.com	4033	6777	2744	40,48%	160	478	318	66,52%
www.ehow.com	16776	28467	11691	41,06%	923	1239	316	25,50%
www.brothersoft.com	6916	11025	4109	37,26%	630	944	314	33,26%
nl.wikipedia.org	360	485	125	25,77%	124	428	304	71,02%
www.hotcourses.com	2177	4741	2564	54,08%	153	455	302	66,37%
hubpages.com	6569	13691	7122	52,01%	157	457	300	65,64%
www.whosdatedwho.com	1438	3699	2261	61,12%	26	323	297	91,95%
www.twenga.co.uk	7015	11447	4432	38,71%	18	305	287	94,09%
www.suite101.com	5571	17430	11859	68,03%	60	342	282	82,45%
directrooms.com	2840	4710	1870	39,70%	149	428	279	65,18%
ezinearticles.com	7106	13644	6538	47,91%	82	354	272	76,83%
www.buzzle.com	4092	9198	5106	55,51%	281	550	269	48,90%
www.1job.co.uk	865	2140	1275	59,57%	26	283	257	90,81%
www.connexions-direct.com	279	1917	1638	85,44%	30	282	252	89,36%

On peut remarquer que, bien que l'échantillon utilisé soit probablement différent de celui de la société Sistrix, dont l'étude a été largement diffusée sur le net, on retrouve grosso modo la même liste de sites et des volumes de perte de visibilité très proches pour ces sites.

Dans cette liste, la typologie des sites impactés est claire (la liste qui suit n'est pas donnée par ordre d'importance) :

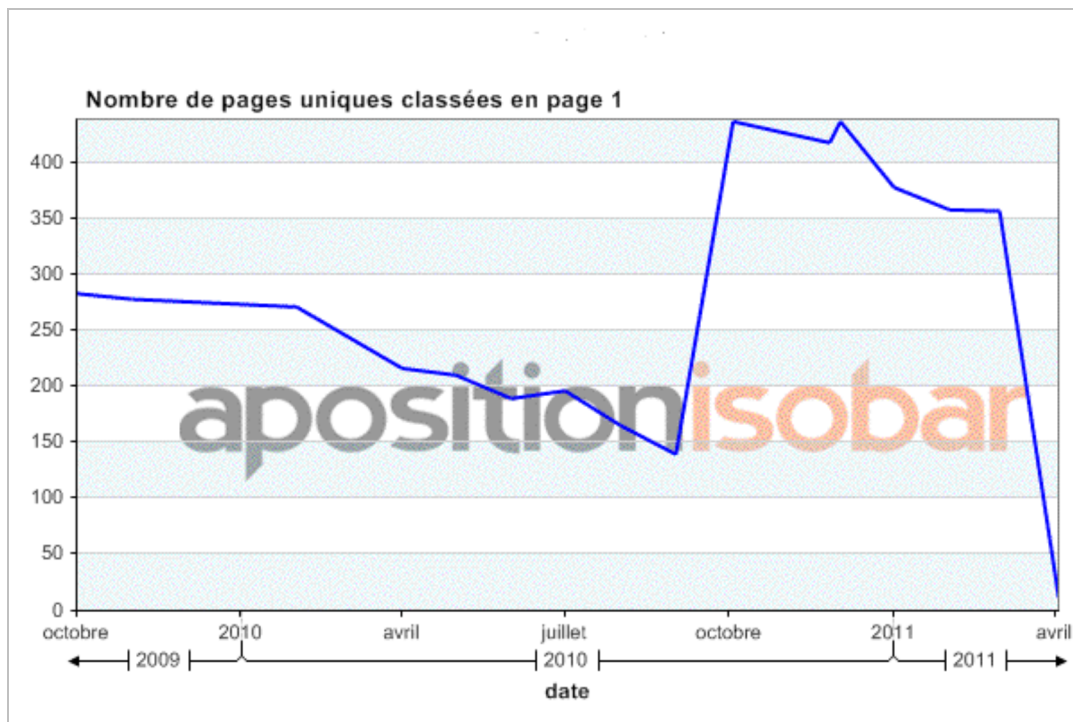
1°) Les annuaires de commerces ou BtoB

On trouve dans cette liste un grand nombre de sites "agrégant" des informations sur les commerces, comme Qype, Hotfrog, Cylex-UK, Londontown, Information-britain, Aboutbritain. Les deux derniers sont des sites remplis de contenu de mauvaise qualité et bourrés de pubs, mais c'est moins vrai pour Qype dont le seul tort est visiblement de créer un grand nombre de pages avec une base de données d'informations sur les commerces collectée sur le web.

2°) Les comparateurs de prix

Il s'agit ici des sites de comparaison de produits : kelkoo, shopzilla, twenga, dooyoo...

Attention : le web britannique regorge de sites de comparaison de services puissants (moneysupermarket.com, comparethemarket.com, gocompare...). Ces sites semblent avoir été moins affectés que les sites de comparaison de produits.



L'impact de Panda sur la visibilité d'un comparateur de prix britannique : le site a pratiquement disparu des premières pages de Google

3°) Les sites d'avis de consommateurs

Ils sont nombreux à avoir été durement touchés : Ciao, Reviewcentre, Dooyoo (qui est également un comparateur de prix)

4°) Les sites de questions réponses

Les sites de questions/réponses construits sur le modèle de Yahoo! Answers ou de Stackoverflow étaient désignés comme des cibles potentielles pour Panda. En tout cas ceux qui ont "pêché" en créant des pages de questions avant d'avoir les réponses, ou du contenu sur des produits réalisé à partir de bases de données destiné à recueillir les FAQ des utilisateurs.

Cela se confirme, on trouve au moins un site de ce type dans la liste : Wisegeek. Mais ces sites sont moins visibles en Grande Bretagne qu'aux USA, on ne les trouve donc pas beaucoup dans le "top" des sites impactés.

5°) Les fermes de contenu

Cette fois-ci, les sites de Demand Media ont été clairement touchés par l'update, notamment eHow (version UK). Mais on trouve aussi d'autres fermes de contenu comme Buzzle, eZinearticles, Suite101 ...

6°) Les sites d'annonces

Les sites d'annonces semblent relativement épargnés par Panda, sauf certains sites d'annonces gratuites de qualité moyenne, comme Homesonsale.

7°) Les agrégateurs d'annonces

Les agrégateurs d'annonces ont été visiblement impactés, mais apparemment moins que d'autres sites. Pourtant certains semblent bien remplir les critères négatifs censés vous précipiter dans l'enfer de Panda. On trouve néanmoins des agrégateurs d'annonce emploi dans la liste (Careerjet, Jobs1, jobs.trovit) et quelques agrégateurs d'annonces immobilières (Look4aproperty).

8°) Les agrégateurs de coupons

Apparemment, la cible n'est pas ici les sites récoltant et diffusant eux-mêmes des réductions chez les commerçants (comme dans le modèle Groupon) mais des sites agrégeant les coupons présents sur les sites marchands. On trouve ainsi dans les sites touchés Myvouchercode, ou Netvouchercode.

9°) Les moteurs de recherche de formation

Les sites collectant sur le net des informations sur les cours et les formations, de type Emagister, figurent également sur la liste.

Notez-bien que cette typologie de sites n'est pas exhaustive.

Quels sont les points communs de ces sites ?

Deux grandes familles de sites se dégagent de cette typologie :

- Des sites dont le modèle de génération de trafic repose essentiellement sur le SEO, et l'utilisation d'une base de données "scrapée" sur le web pour créer des pages.
- Ou des sites dont le contenu peut-être original, mais considéré comme "de mauvaise qualité", et dont le modèle repose également sur la création de trafic via le SEO.

Il est difficile de savoir à ce stade si les dix familles de sites identifiées plus haut sont particulièrement ciblées, autrement dit si ces sites ont fait l'objet d'un profilage pour les reconnaître, et les éliminer des SERPs, ou si l'on assiste à un effet massif dû à la prise en compte de critères d'usage révélant des sites "de mauvaise qualité".

On peut constater, non sans malice, que les nouveaux critères de "qualité" choisis par Google ont tendance à éliminer des pages de résultats des sites dont le modèle "menace" les nouveaux modèles économiques de Google :

- Compérateurs de prix, concurrents de Google Products ;
- Sites de coupons, concurrent du nouveau service de Google (Google Offers) lancé aux USA ;
- Annuaire locaux, concurrents de Google Places et sur la publicité locale ;
- Agrégateurs divers et variés, par exemple dans le domaine du tourisme ;
- Et d'une manière générale, tout site qui s'interpose entre l'utilisateur et un fournisseur de produits, de service, ou d'information.

Le fait que les sites organisent mieux l'information que le moteur de recherche, ou apportent une valeur ajoutée, ne semble clairement pas suffire à échapper aux foudres du Dieu Google ! L'élément nouveau, c'est que l'impact de Panda est tel pour ces sites que l'équilibre économique des entreprises concernées peut-être durablement fragilisé. On peut donc s'attendre à une multiplication des recours contre Google pour pratiques anticoncurrentielles. La foudre est par exemple tombée une deuxième fois sur Ciao, le site d'avis de Microsoft, qui avait déjà intenté un recours contre Google pour ce motif.

Quels sites ont le plus bénéficié de Panda sur l'index UK ?

Comme toujours avec les changements dans l'algorithme, si certains disparaissent des premières pages, d'autres en bénéficient forcément !

Lorsque l'on regarde qui figure dans le top des sites ayant gagné fortement en visibilité avec Panda, on voit clairement que cette mise à jour renforce encore la position des principaux sites marchands et poids lourds du web. En fait, la part de voix des "petits" sites marchands, ou d'information s'est sans doute réduite, même si l'impact à court terme de ces changements est forcément positif pour les sites non impactés.

Croissance pour les principaux sites d'information ou de contenu :

Domaine	Progression en visibilité sur les premières pages de Google
www.imdb.com	9,5%
www.guardian.co.uk	14,7%
www.telegraph.co.uk	20,9%
www.dailymail.co.uk	22,9%
www.merriam-webster.com	160,0%

Croissance pour les principaux sites e-commerce :

Site	Progression en visibilité sur les premières pages de Google
www.amazon.co.uk	18,7%
shop.ebay.co.uk	28,4%
www.laterooms.co.uk	33,8%
www.argos.co.uk	34,7%

On trouve également dans les grands gagnants de l'update Panda, l'annuaire des pages jaunes britannique, Yell (+34,7%), le plus important sites d'annonces gratuites local, Gumtree (+68% !) et Which.co.uk, un site social d'avis d'experts et de tests, (+53%)...

L'impact de Panda sur les résultats de Google :

Pour comprendre ce qui s'est réellement passé lors de la mise à jour, le plus parlant est sans doute de comparer les positions pré et post panda sur deux requêtes qui ont été violemment chamboulées par l'update.

Requête : *bagless vacuum cleaner*

Position	pre Panda	Evolution	post Panda
1	www.argos.co.uk	-	www.argos.co.uk

2	www.argos.co.uk	-	www.argos.co.uk
3	reviews.argos.co.uk	(collapsé : résultat clusterisé)	www.which.co.uk
4	www.pixmania.co.uk	disparu des 100 premiers résultats	www.comet.co.uk
5	www.pixmania.co.uk	disparu des 100 premiers résultats	www.currys.co.uk
6	www.reviewcentre.com	passé en page 3 (pos 23)	www.currys.co.uk
7	www.which.co.uk	->pos 3	www.hoover.co.uk
8	www.which.co.uk		www.hoover.co.uk
9	www.comet.co.uk	->pos 4	www.hoover.co.uk
10	www.kelkoo.co.uk	passé en page 6 (pos 53)	direct.tesco.com

Après le changement, les sites d'avis, les marketplaces et les comparateurs de prix ont disparu de la première page. Il ne reste plus que des sites marchands et un site de tests et d'avis !

Requête : chanel no 5 body lotion

Position	Pre Panda	evolution	Post Panda
1	www.comparestoreprices.co.uk	8 places perdues -> pos 9	www.chanel.com
2	www.comparestoreprices.co.uk	idem	www.chanel.com
3	www.ciao.co.uk	->page 2, pos 13	www.chanel.com
4	www.ciao.co.uk	idem	www.forget-flowers.co.uk
5	www.ciao.co.uk	idem	health-beauty.shop.ebay.co.uk
6	www.shopzilla.co.uk	->page 2, pos 13	cgi.ebay.co.uk
7	www.shopzilla.co.uk		cgi.ebay.co.uk
8	www.shopzilla.co.uk		cgi.ebay.co.uk
9	www.dealtime.co.uk	page 2, pos 11	www.comparestoreprices.co.uk
10	www.dealtime.co.uk		www.gumtree.com
11	www.chanel.com	+10, pos 1	www.dealtime.co.uk
12	www.chanel.com		www.shopzilla.co.uk
13	www.chanel.com		www.ciao.co.uk
14	health-beauty.shop.ebay.co.uk		en.wikipedia.org
15	shop.ebay.co.uk		www.mistrys.co.uk

Cette requête est intéressante car beaucoup de sites visés par Panda la ciblent. La première page était "squattée" par les comparateurs de prix et par les sites d'avis de consommateurs. Après Panda, le site Chanel remonte en première position, et les sites d'eBay gagnent dix places sur cette requête...

Quelles conclusions en tirer ?

Panda est la première mise à jour depuis longtemps qui chamboule autant les classements, et surtout, pénalise autant de sites d'un coup. L'impact de Panda sur un site peut être dévastateur : le site est évalué dans sa globalité, ce qui signifie que si un site est jugé de mauvaise qualité, même ses meilleures pages peuvent être déclassées.

Jusqu'ici, les pénalités automatiques infligées par Google étaient plus limitées, au point que leurs effets n'étaient pas toujours détectés par les webmasters. Les pénalités "manuelles" ou "supervisées par un humain" concernaient un nombre limité de sites, ce qui permettait clairement à de nombreux autres de continuer impunément à violer les TOS de Google.

Cette mise à jour de l'algorithme sonne comme le début d'une nouvelle ère pour le SEO. Le service qualité de recherche de Matt Cutts, débordé depuis la mise en place de Caffeine par la quantité de nouvelles pages relevant du spamdexing, est maintenant secondé par une évaluation automatique de la qualité des pages. Les critères utilisés, essentiellement des critères d'usage, de qualité, d'autorité, empêchent durablement de bâtir une stratégie à long terme de développement du trafic basée uniquement sur le SEO.

Le fait que Google ait décidé de frapper à grande échelle des sites piliers d'écosystèmes bien établi de la netéconomie est également quelque chose de très nouveau. Son discours des origines, exprimant une neutralité bienveillante vis-à-vis d'un web sur lequel son influence se limitait à organiser l'information, est bien mort et enterré. Google est un acteur majeur de tout l'écosystème, et n'hésite plus à modifier les règles du jeu, puisqu'il en a le pouvoir, surtout si les acteurs vivent comme des parasites à ses dépens, ou, plus simplement, ont pris le risque de choisir un modèle économique qui les rendent dépendants de Google.

Quel impact attendre de Panda en France ?

A la date de rédaction de ces lignes, aucun élément ne permettait de prévoir la date du déploiement en France. Une mise à jour comme Panda nécessite apparemment d'adapter les paramètres de l'algorithme à chaque langue et même peut-être à chaque index pays.

Habituellement, quelques semaines séparent les étapes d'un déploiement : la prochaine étape est sans doute pour bientôt. Mais on ne sait pas si l'algorithme sera déployé dans toutes les autres langues ou pays d'un coup, ou si l'on passera d'abord par une mise à jour en allemand ou en espagnol d'abord.

Matt Cutts a laissé entendre lors d'une conférence qu'il s'attendait à ce que le nombre de requêtes impactées par Panda soit beaucoup plus réduit dans d'autres langues. Cela signifie-t'il que le Panda français sera moins "agressif" que le Panda anglais ? Probablement pas. Cette prédiction se base sur un constat : les formes de spamdexing ciblées par l'algorithme étaient simplement beaucoup plus fréquentes et gênantes sur les sites en langue anglaise. Statistiquement, les classements risquent d'être moins modifiés en France.

Mais cela ne signifie pas que Panda ne frappera pas durement des sites des catégories citées plus haut dans cet article. Il ne reste que quelques jours (ou semaines...) pour essayer d'améliorer les critères de qualité de vos sites.

Le problème, c'est que ces critères sont déjà peut être en cours d'évaluation. Et qu'il est compliqué de les faire bouger du jour au lendemain de manière notable : une notoriété, une réputation, une relation privilégiée avec ses utilisateurs, ne se crée pas en trois semaines. Pour certains, il est peut-être déjà trop tard...

Philippe YONNET, Directeur SEO international, Twenga et **Jérôme SPIRAL**, Directeur Stratégie & Développement, Aposition - Isobar

Réagissez à cet article sur le blog des abonnés d'Abondance :
<http://blog-abonnes.abondance.com/2011/05/panda-analyse-des-effets-de-la-mise.html>