

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

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

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

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

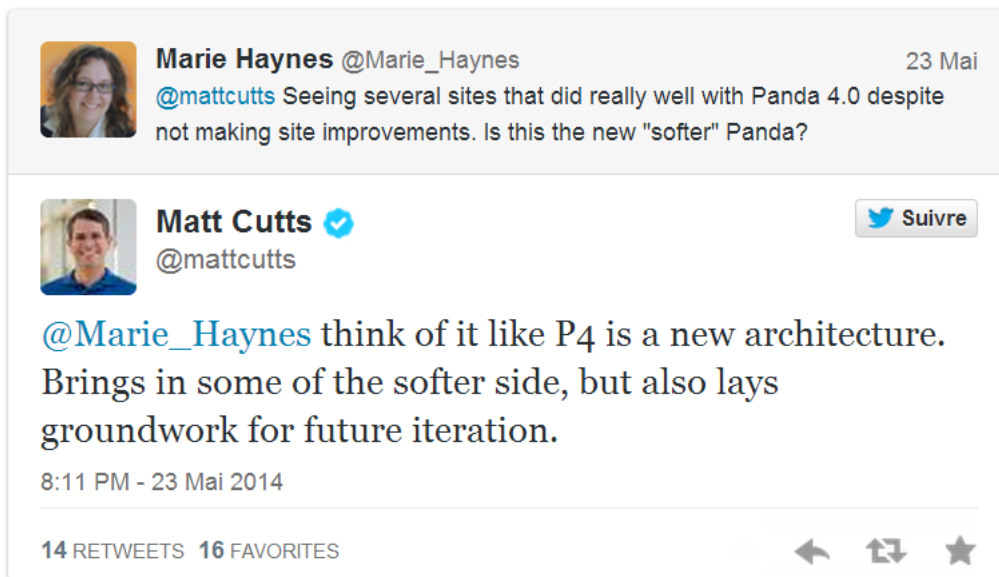
Panda 4.0 : une mise à jour atypique

Une mise à jour "officielle"

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



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



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

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

Déploiement de filtres en série

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

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

Une mise à jour plus "douce" ?

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

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

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

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

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

Les perdants et les gagnants sur GOOGLE USA

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

Google SERPs: Panda 4.0 Losers

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

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

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

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

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

Google SERPs: Panda 4.0 Winners

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

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

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

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

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

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

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

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



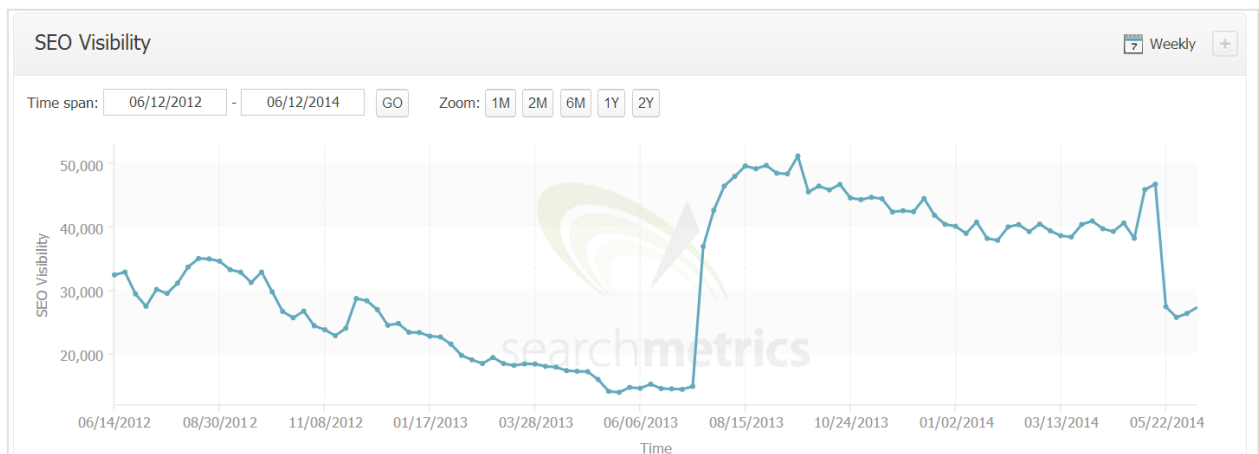
Pour cet exemple, on peut penser qu'un changement de critères est à l'oeuvre, car fondamentalement le site n'a pas changé de contenu à ce point.

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



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

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



Quoi de neuf dans cette mise à jour alors ?

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

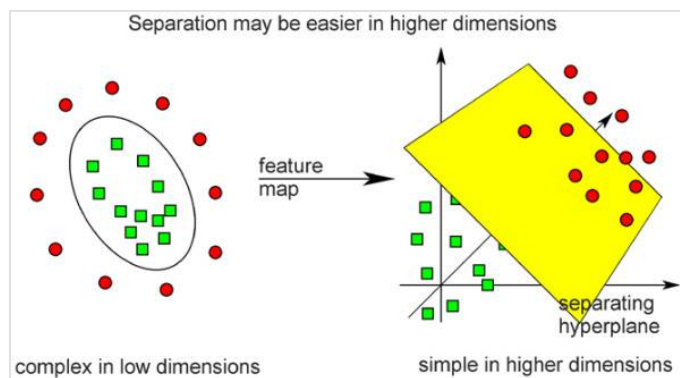
Mais pour mieux comprendre ce qui a changé, il faut faire un retour en arrière et faire le point sur ce que l'on sait sur le filtre Panda.

Panda, un "classifieur" de sites

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

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

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



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

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

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

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

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

Le brevet de Navneet Panda

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

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

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

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

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

La mauvaise qualité des liens

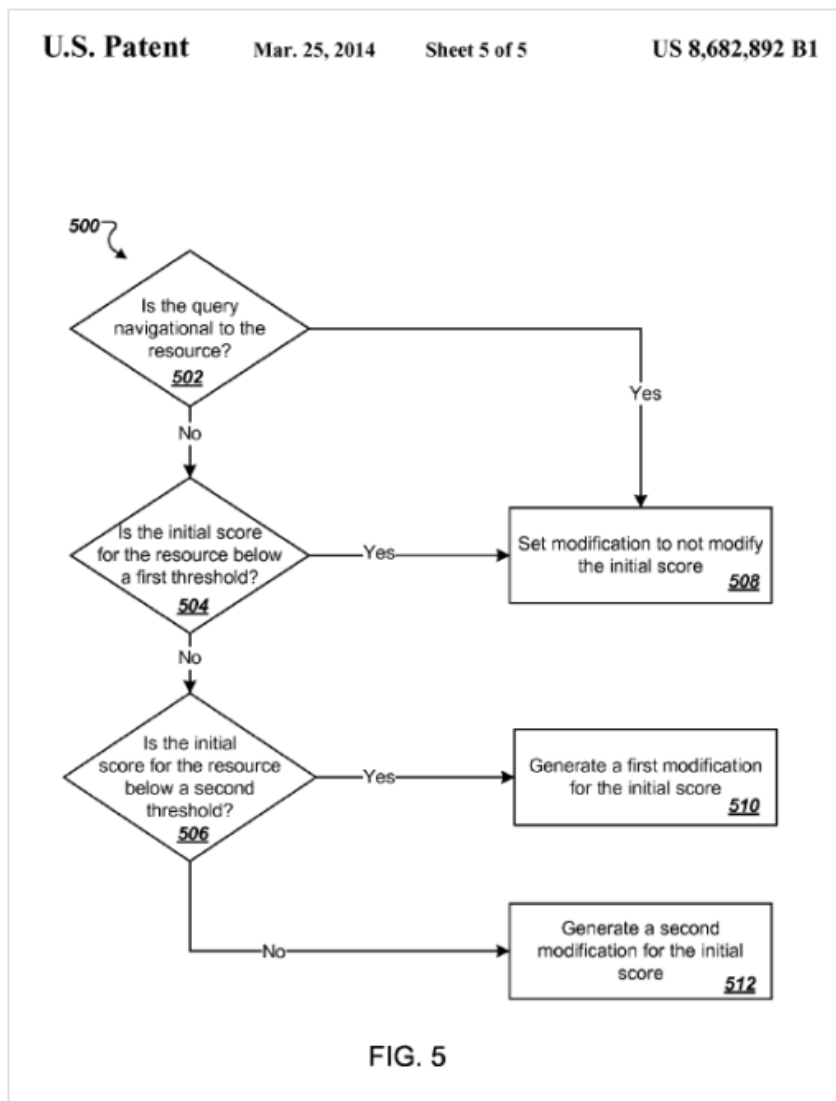
Le brevet décrit une méthode permettant de filtrer, parmi les backlinks, tous les liens que contrôle le webmaster, pour créer à la fin un indicateur dépendant du nombre de liens « indépendants » (non manipulés, donc indépendants vis-à-vis du webmaster/éditeur du site). Un grand nombre de liens indépendants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

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

Les requêtes de référence

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

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



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

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

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

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

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

Bibliographie

Lien vers l'article du blog de Searchmetrics :

<http://blog.searchmetrics.com/us/2014/05/21/panda-update-4-0-winners-and-losers-google-usa/>

Lien vers l'article du blog de search-foresight :

<http://www.search-foresight.com/panda-4-0-les-gagnants-les-perdants-en-france/>

Le brevet de Navneet Panda et Vladimir Ofitserov :

Ranking search results

<http://patft.uspto.gov/netacgi/nph->

[Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetacgi%2FPTO%2Fsearch-adv.htm&r=1&f=G&l=50&d=PALL&S1=08682892&OS=PN/08682892&RS=PN/08682892](http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetacgi%2FPTO%2Fsearch-adv.htm&r=1&f=G&l=50&d=PALL&S1=08682892&OS=PN/08682892&RS=PN/08682892)

Le post d'annonce de Google Panda en février 2011 :

<http://googleblog.blogspot.com/2011/02/finding-more-high-quality-sites-in.html>

Le billet de Bill Slawski à propos du brevet de Navneet Panda :

<http://www.seobythesea.com/2014/03/googles-panda-granted-patent-on-ranking-search-results/>

Philippe YONNET , *Directeur Général de l'agence Search-Foresight*
(<http://www.search-foresight.com>).