

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

La grande affaire de ce mois de mars 2011 en termes de référencement naturel aura été la mise à jour de l'algorithme de Google nommée "Farmer" par les uns, "Panda" par les autres et qui vise en particulier les sites proposant des contenus de faible qualité spammant l'index du moteur de recherche. Quel est l'historique de cette mise à jour ? Pourquoi a-t-elle été imaginée ? Comment savoir quand elle sera mise en place en France ? Comment faire en sorte que votre contenu ne soit pas touché par cette nouveauté ? Voici quelques pistes de réflexion pour éviter un coup de "Kung Fu Panda"...

Ce début d'année aura été marqué par un important changement d'algorithme de la part du moteur de recherche leader, dont les effets dévastateurs ne sont pas sans rappeler les anciennes Google Dance(s). Qu'est-ce que Google Panda, et comment s'en prémunir ? Les spéculations vont bon train sur le Web...

Rappel des faits

Fin 2010, Matt Cutts laisse entendre que la lutte contre le spam sera cause prioritaire en 2011 pour Google. Le Cloaking est en ligne de mire (voir <http://searchengineland.com/google-vows-to-look-at-deceptive-cloaking-techniques-59802>). Il faut dire que Google est de plus en plus attaqué sur la qualité de ses résultats de recherche !

Début 2011, les premiers changements se font sentir dans Google. Dans un article (<http://googleblog.blogspot.com/2011/01/google-search-and-search-engine-spam.html>), Matt Cutts explique qu'un nouveau système de classement est mis en place, et qu'il est capable de détecter le contenu spammant de type "succession de mots clés". Il explique aussi que Google va prochainement faire quelque chose contre les fermes de contenu.... La menace se précise...

Le 28 janvier 2011, Matt Cutts annonce une première mise à jour de l'algorithme de classement (<http://www.mattcutts.com/blog/algorithm-change-launched/>). Destinée aux sites qui volent le contenu d'autres sites (les "scrapers"), cette mise à jour "mineure" n'impacte que 2% des résultats de recherche selon Matt Cutts.

Un mois tard, le 24 février, Google sort cette fois l'artillerie lourde (<http://googleblog.blogspot.com/2011/02/finding-more-high-quality-sites-in.html>). Cette mise à jour, qui s'en prend aux "sites de faible qualité", va impacter 12% des résultats de recherche. Elle est baptisée "**Farmer**" par Danny Sullivan du site SearchEngineLand, car elle cible particulièrement les fermes de contenu (<http://searchengineland.com/google-forecloses-on-content-farms-with-farmer-algorithm-update-66071>) mais le nom officiel sera **Panda** (du surnom d'un des développeurs ayant travaillé sur le sujet chez Google). A noter que dans le même temps, Google lance une extension Chrome permettant de signaler les sites spammant (<http://actu.abondance.com/2011/02/une-extension-chrome-pour-bloquer-les.html>).

La mise à jour Panda/Farmer doit prochainement se déployer en France (si ce n'est déjà fait au moment où vous lirez ces lignes : <http://actu.abondance.com/2011/03/google-panda-plus-dinfos-et-arrivee-en.html>) après la Grande-Bretagne et les requêtes anglophones en général : <http://actu.abondance.com/2011/04/google-panda-uk-nouvelles-stats-sur-les.html>. C'est donc le moment idéal pour faire le point...

Kung Fu Panda

Les deux ingénieurs Google, Amit Singhal et Matt Cutts, se sont exprimé lors du salon TED 2011 sur cette mise à jour de l'algorithme (<http://www.wired.com/epicenter/2011/03/the-panda-that-hates-farms/all/1>). Outre que le nom "officiel" de la mise à jour a été confirmé comme étant Panda, on apprend que la mise à jour Caffeine (courant 2009) a beaucoup joué

sur la taille de l'index Google et sur la fraîcheur des résultats. Le problème était que beaucoup de nouvelles pages indexées n'étaient pas forcément très pertinentes et que Google a donc indexé une grande quantité de Spam suite à ce changement de structure de l'index.

Il a donc fallu trouver un moyen de détecter le contenu peu pertinent ("shallow-content") : pour cela les ingénieurs ont pris en compte la grille de validation utilisée par les correcteurs humains (Google utilise en effet des testeurs humains, appelés "quality raters", pour vérifier la qualité des sites).



Une fois qu'il a été trouvé une définition claire de ce qu'était un site de faible qualité (définition élaborée également en analysant la façon dont les internautes bloquaient les sites dans Chrome), un algorithme basé sur plusieurs signaux a été élaboré : Google Panda était né !

Un peu plus tard, Matt Cutts s'est à nouveau exprimé au salon SMX West en compagnie de Maile Ohye (voir <http://searchengineland.com/lessons-learned-at-smx-west-googles-farmerpanda-update-white-hat-cloaking-and-link-building-67838>). Selon lui, la mise à jour est entièrement automatisée, il n'y a aucune intervention humaine sur les résultats de recherche. On y apprend que les pénalités sont mises en place de façon globale, sur un site entier plutôt que sur une page. Maile Ohye déclare d'ailleurs qu'il vaut mieux bloquer l'indexation d'une page (par une balise meta "robots") tant que le contenu n'est pas de bonne qualité ! Matt Cutts indique également à cette occasion que Google diminue sa fréquence d'indexation s'il juge qu'un site est peu pertinent.

Nouvelles règles pour le SEO

Des maigres éléments fournis par Google (comme d'habitude, la recette de l'algorithme est gardée secrète) et des retours d'expérience des webmasters, on peut retirer quelques éléments intéressants :

Eviter le Cloacking sous toutes ses formes

Le Cloacking (présentation d'un contenu différent aux internautes et aux moteurs) fait partie des cibles de Google Panda. Pour Google, les règles sont durcies : il n'y a pas de "bon Cloacking" et de "mauvais Cloacking".

Les techniques suivantes sont donc sanctionnées :

- * Proposer un contenu HTML aux moteurs et un contenu Flash aux internautes.
 - * La réécriture d'URL uniquement à l'attention des moteurs de recherche.
 - * Proposer une image aux internautes ou du texte aux moteurs (texte caché derrière l'image par un style CSS approprié).
 - * Montrer un contenu différent aux nouveaux visiteurs et aux anciens visiteurs d'un site
- D'une manière générale, il faut donc éviter de présenter un contenu différent suivant le visiteur. Le mieux est de donner le choix aux internautes, comme de visiter ou pas la version Flash d'un site.

Tu ne voleras point

La première mise à jour de l'année à porté sur les "Scrapers", ces sites qui récupèrent du contenu à droite à gauche pour créer artificiellement des pages de contenu, qui peuvent sembler pertinentes de prime abord.

Google semble désormais capable de détecter les contenus copiés depuis d'autres sites, et la plus grande prudence doit être de rigueur.

Si vous utilisez un contenu qui provient d'un site tiers, il faut le faire en toute transparence : citez vos sources !

Il peut aussi être intéressant d'utiliser la balise de syndication proposée par Google News : cette balise permet d'indiquer à Google quelle est la source d'un article (voir <http://googlenewsblog.blogspot.com/2010/11/credit-where-credit-is-due.html>)

Exemple d'une telle balise :

```
<meta name="syndication-source" content="http://www.publisherX.com/wire_story_1.html">
```

Si la duplication de contenu concerne des sites dont vous êtes propriétaire, vous pouvez tenter d'utiliser la balise `rel=canonical` préconisée par Google. Même si cette balise reste une indication pour Google, elle pourra servir à montrer sa bonne volonté et à indiquer quel site doit être mis en avant dans les résultats de recherche :

```
<link rel="canonical" href="http://www.monsite1.fr">
```

Article marketing : à éviter ?

Créer un contenu original et "pertinent" pour Google est un point essentiel pour passer à travers l'algorithme Panda. Reste à savoir ce que Google appelle un contenu pertinent : Panda est capable de détecter les fermes de contenu, donc la détection est assez poussée.

Une étude de SEOMOZ montre que l'utilisation de contenu généré par les utilisateurs (commentaires et avis) est une solution appréciée par Panda (voir

<http://www.seomoz.org/blog/ugc-gets-an-a-on-google-test-with-panda-update-12260>). Il suffirait donc juste de laisser les internautes s'exprimer et ajouter du contenu sur une page produit ?

Ce n'est donc pas forcément la qualité rédactionnelle qui prime, mais plutôt l'originalité et la pertinence du contenu.

Dans une vidéo publiée sur Youtube (<http://www.youtube.com/watch?v=x5xP-pTmlpY>), Matt Cutts donne son avis les "articles marketing", c'est-à-dire les articles qui sont créés uniquement pour générer des liens et publiés à l'identique sur des nombreux portails peu qualitatifs. Pour lui, ce n'est pas pertinent pour le référencement et cela surcharge le web de contenus peu pertinents.

Les articles marketing ne sont pas morts, mais il faut miser sur des contenus originaux et susceptibles d'intéresser les internautes, et choisir soigneusement les sites de diffusion (éviter par exemple les portails qui acceptent des textes sans aucune condition). Privilégier la qualité plutôt que la quantité !

Le contenu personnalisé est la clé

L'idéal pour éviter le "coup de Panda" est évidemment de proposer un contenu pertinent et personnalisé sur chaque page de son site, mais tous les webmasters savent bien que c'est très difficile à mettre en œuvre ! Page de mentions légales, conditions de vente, résultats de recherche, archives des blogs... Il existe des quantités de pages utiles aux internautes mais peu "Panda Friendly".

La solution, conseillée par Maile Ohye (en charge du centre d'aide Google) est d'empêcher l'indexation de pages non pertinentes en termes de contenu.

Pour cela, la solution préconisée est d'inclure une balise `<meta name="robots" content="noindex, follow">` sur les pages présentant du contenu peu qualitatif...

On se souvient de la mode du PageRank sculpting qui consistait à bloquer ou pas certains liens afin de transférer au mieux le PageRank. Voici venu le temps du **Content Sculpting** :

empêcher Google d'accéder à des pages de contenu peu pertinentes, afin de réussir l'examen de l'algorithme Panda...

Améliorer le temps de réponse de son site

Cela fait déjà quelques temps que Google tient compte du temps de réponse d'une page web dans son algorithme. Matt Cutts a confirmé au salon SMX que c'était un facteur important pour le positionnement. En effet, si Google détermine qu'un site est trop lent à utiliser pour les internautes, il le déclassera automatiquement.

On ne saurait trop conseiller l'utilisation de l'outil de diagnostic PageSpeed (imaginé par Google), qui vient justement de se doter d'une version en ligne : <http://pagespeed.googlelabs.com/>

Trop de pub tue le contenu

Restant assez évasif sur la question, Matt Cutts a affirmé que l'utilisation de publicité sur un site n'était pas en soi un motif de sanction mais qu'une surabondance de pub constituera un signal négatif et un motif de sanction.

Il va de soi que Google ne va pas interdire les AdSenses sur un site web ! Néanmoins il faut garder à l'esprit qu'il faut conserver un bon ratio entre la publicité et le contenu pertinent.

Faire le buzz et obtenir des liens naturels

Matt Cutts continue d'affirmer qu'un bon site, ayant du contenu pertinent, obtiendra naturellement des liens qualifiés. Plus facile à dire qu'à faire !

Les liens "artificiels", de type échanges de liens entre sites qui n'ont aucune activité en commun, est peu apprécié par Google.

Comme nous l'avons vu précédemment, les articles marketing et communiqués de presse n'ont plus également beaucoup la cote pour Google, il vaut donc mieux se faire connaître auprès des internautes et des sites faisant autorité.

Pour cela, l'utilisation de Twitter et de Facebook semble être un bon moyen, car Google considère favorablement les liens issus de ces réseaux sociaux. Le SMO est promis à un bel avenir !

Conclusion : Google Panda bientôt en France ?

Dans un article posté le 11 avril (<http://googlewebmastercentral.blogspot.com/2011/04/high-quality-sites-algorithm-goes.html>), Amit Singhal affirme que Panda est déployé pour tous les utilisateurs de langue anglaise, et que le lancement se fera progressivement dans les autres langues. Officiellement Panda n'est donc pas arrivé en France, et pourtant... Certains webmasters ont déjà constaté des chutes radicales de positionnement à la mi-avril.

Les symptômes de pénalisation seront certainement une chute brutale de positions et de la vitesse d'indexation de Google : ce dernier point peut être mis en valeur à l'aide de Google Analytics (voir la méthodologie proposée sur <http://www.wordtracker.com/academy/google-panda-farmer>).

Si votre site web n'est pas victime de ces symptômes, il peut être néanmoins utile de réfléchir dès maintenant à la qualité rédactionnelle de votre contenu : bloquer les pages non pertinentes, supprimer toute technique pouvant être assimilable à du cloacking, ajouter des champs commentaires pour développer le contenu généré par les internautes sont des actions relativement simples à mettre en œuvre.

L'idéal est évidemment de créer du contenu original et intéressant mais on n'a pas toujours le temps et les moyens de le faire! Quoiqu'il en soit, les référenceurs français attendent Google Panda de pied ferme et sont prêt à lui résister...

Guillaume Thavaud, cellule de veille de la société Brioude-Internet Référencement
(<http://www.brioude-internet.fr>).

Réagissez à cet article sur le blog des abonnés d'Abondance :

<http://blog-abonnes.abondance.com/2011/04/google-pandafarmer-la-chasse-au-spam.html>