

## Vince : une mise à jour réellement majeure de l'algorithme de Google ?

[Retour au sommaire de la lettre](#)

<b>Domaine :</b>	Recherche	<b>Référencement</b>
<b>Niveau :</b>	Pour tous	<b>Avancé</b>

*Les forums spécialisés ont largement parlé dernièrement de Vince, une mise à jour de l'algorithme de pertinence de Google (du nom de l'un des ingénieurs ayant travaillé dessus) qui a fait couler beaucoup d'encre virtuelle... Qu'en est-il réellement ? Ce changement affecte-t-il et fait-il la part belle, comme il a été dit, aux grandes marques ? Pour en avoir le cœur net, voici quelques tests qui devraient grandement vous aider à vous faire votre opinion sur la question, et à appréhender quelques futures voies d'amélioration que le moteur de recherche leader est en train d'explorer...*

La mise à jour "Vince" est l'un des derniers "buzz" qui s'est répandu dans la communauté du référencement. Les forums et les blogs bruissent de rumeurs autour d'un changement dans l'algorithme de Google qui tendrait à favoriser les marques dans les classements, selon la plupart des commentateurs.

Tout à débuté par un sujet lancé sur Webmasterworld.com, comme souvent dans l'histoire de la détection des "updates" - "mises à jour" - de l'algorithme de classement de Google. Dans un message daté du 20 février 2009 (Is Google Giving more Offline Brand Authority in SERPS ? <http://www.webmasterworld.com/google/3854121.htm>), l'un des membres s'étonnait de voir de plus en plus de sites de grandes marques apparaître sur des requêtes comme [Laptop](#), [High Speed Internet](#), [Quit Smoking](#), [Car Audio](#)... Ses remarques n'ont pas rencontré un grand écho jusqu'à ce que Aaron Wall (l'éditeur du site SEO Book) annonce le 25 février qu'il avait remarqué la même chose et qu'il publiait un article sur le sujet. Cet article (<http://www.seobook.com/google-branding>) a rapidement fait le tour de la communauté des référenceurs anglo-saxons, suscitant toutefois autant de commentaires sceptiques que de confirmations enthousiastes.

Mais Matt Cutts le responsable de l'équipe "search quality" chez Google a confirmé qu'il y'avait bien eu changement dans l'algorithme dans une video réalisée le 26 février et diffusée le 4 mars sur Youtube (<http://www.youtube.com/watch?v=LMfWPWUh5uU>).



Matt a également révélé que ce changement avait un nom de code chez Google : "Vince", du nom de la personne qui a travaillé sur le dossier.

Dès lors que ce changement dans l'algorithme est avéré, plusieurs questions viennent immédiatement à l'esprit :

- Quand s'est déroulée cette "mise à jour" ? Car la plupart des observateurs n'avaient rien détecté...
- Pourquoi cette mise à jour est elle passée inaperçue ?
- Comment interpréter les commentaires de Matt Cutts sur le sujet ?

- Est-ce que oui ou non, il y'a un rapport avec la notion de "marque" ?
- Quelles sont les hypothèses que l'on peut formuler pour expliquer les changements dans les pages de résultats ?

### **Quand cette mise à jour a-t-elle eu lieu ?**

Si on regarde la chronologie des remontées d'information sur Vince, on se rend compte que les premières observations remontent au 20 février 2009, mais que la plupart des commentaires font remonter les changements à... plusieurs mois. Aaron Wall a quant à lui donné une date précise en s'appuyant sur les données fournies par l'outil Rankpulse (<http://www.rankpulse.com>).

L'outil Rankpulse interroge Google quotidiennement sur 1 000 requêtes et permet de suivre grâce à des graphes l'évolution des pages de résultats jour après jour. Par exemple, si on regarde la requête "airline tickets" à la date du 5 avril 2009, on obtient les positions suivantes :

- #1 **www.travelocity.com**
- #2 www.orbitz.com
- #3 www.priceline.com
- #4 **www.nwa.com**
- #5 www.cheaptickets.com
- #6 supersearch.travelzoo.com
- #7 www.kayak.com
- #8 **www.aa.com**
- #9 **www.delta.com**
- #10 **www.continental.com**

Notons que l'outil rankpulse ne montre que les domaines, et non les pages - or ce sont bien les pages qui obtiennent des classements - et qu'il s'appuie sur l'API de Google et donc fournit des résultats souvent différents de ceux que voient les internautes.

Les domaines en rouge sont ceux qui ont fait "irruption" dans la première page de résultats, tous ensemble, le 18 janvier 2009. La page du site continental.com a également été violemment mise en avant à la même date, mais n'est entrée sur la première page que le 1er février.

Sur cette requête, ce sont donc quatre compagnies aériennes qui montent brutalement en tête des résultats (Northwest, American Airlines, Delta Airlines, et Continental Airlines), plus un important site de vente en ligne de billets d'avion.

Aaron Wall a donné d'autres exemples de requêtes avec des changements visibles le 18 janvier ([watches](#), [online degree](#), [diets](#), [health insurance](#), [boots](#), [autoinsurance](#))

Voici un autre exemple intéressant : la requête "mortgages", toujours étudiée au 5 avril 2009

- #1 www.bankofamerica.com
- #2 www.mortgage101.com
- #3 **www.lendingtree.com**
- #4 www.bankrate.com
- #5 www.bankrate.com
- #6 **www.wellsfargo.com**
- #7 **my.countrywide.com**
- #8 en.wikipedia.org
- #9 www.mortgage.com
- #10 www.mortgagefindersnetwork.com

Les pages en rouge ont acquis leurs positions en première page... le 18 janvier. Ceci montre que l'on peut effectivement trouver d'autres exemples que ceux qu'Aaron Wall a détectés.

Mais si on fait l'exercice avec une requête similaire à celle testée par Voasi (notre très perspicace membre de WMW) : "desktops" ([laptops](#) n'est pas dans la base de rankpulse), on

s'aperçoit effectivement que Dell prend la deuxième place sur cette requête mais vers le 20 février !

Dans le même temps d'autres requêtes apparemment similaires ne montrent pas de changements mi janvier 2009, comme "[car loans](#)" par exemple.

Il faut donc prendre avec prudence les affirmations d'Aaron Wall : il s'est sans doute passé quelque chose le 18 janvier, mais il y a des changements de classement sensibles sur des requêtes concurrentielles très régulièrement. Bref, la date officielle d'introduction de la modification n'est pas vraiment connue.

## **Pourquoi cette mise à jour est-elle passée inaperçue ?**

La mise à jour Vince, si elle a bien eu lieu le 18 janvier, n'a été "détectée" qu'un mois après : pourquoi ? Probablement parce qu'elle n'affecte qu'un nombre limité de requêtes.

C'est exactement ce que dit Matt Cutts dans son commentaire du 26 février :

*"Tout d'abord, OUI, Google a modifié sa manière de classer les pages, c'est l'un des 300 à 400 changements que nous faisons par an. C'est pourquoi je n'appellerais pas cela une "mise à jour", j'appellerais cela un simple changement. Si vous devez trouver une appellation, sachez que l'une des personnes qui a beaucoup travaillé là dessus, son nom est Vince. Ce changement là, nous en parlons comme le "changement Vince" au Googleplex.*

*Je n'appellerais pas cela vraiment une mise à jour, mais il y a eu un au moins une modification dans la manière dont nous réalisons certains classements. Cela n'affecte pas une large majorité de requêtes, c'est plus volontiers... et la plupart des gens ne l'ont pas remarqué. Je veux dire que oui Aaron en a parlé, et je pense que même avant ça des gens sur Webmasterworld en avaient parlé également. Mais cela affecte un nombre relativement réduit de requêtes, ce n'est pas comme si cela affectait une masse de requêtes longue traine ou quelque chose comme ça."*

Ce qui est intéressant dans cette mise à jour, c'est qu'il s'agit apparemment de quelque chose de réel et d'important, mais que cela n'a déclenché aucun signal d'alerte dans les outils que nous utilisons chez Aposition. Ces outils, que ce soient ceux qui surveillent les positions des clients, ou les changements dans les SERPS (*Search Engine Result Page*), donnent des informations statistiques et ces changements semblent totalement noyés dans le bruit des évolutions "normales" des SERPs (*voir plus loin*).

## **Vince favorise-t-il les sites de marque ?**

Les remarques sur la montée de Dell, Des compagnies aériennes comme Northwest ou American Airlines, ou des banques spécialisées dans le crédit comme Wells Fargo ou Countrywide Financial, ont conduit un grand nombre d'observateurs à voir dans ce changement une prise en compte de la puissance des marques comme critère de classement.

Cette opinion a été fortement influencée par des propos qu'Eric Schmidt a tenus en octobre 2008 devant des cadres de la presse magazine qui visitaient le Googleplex dans le cadre de leur conférence annuelle. Le Vice Président de Google leur a expliqué que leurs marques étaient des signaux de plus en plus importants puisqu'elles montrent que l'on peut avoir confiance dans le contenu qu'ils mettent en ligne.

*"L'internet devient rapidement "une fosse d'aisance" où de fausses informations prospèrent, a affirmé Eric Schmidt. Les marques sont la solution, pas le problème. C'est grâce aux marques que vous pouvez filtrer ce qu'il y'a de dans la fosse d'aisance.[...] L'affinité pour les marques est clairement programmée dans nos cerveaux. C'est si fondamental dans l'existence humaine que cela persiste. Il doit avoir une composante génétique là dedans."*

Depuis lors, beaucoup de référenceurs pensent soit que Google intègre déjà la puissance des

marques comme critère dans leur algorithme, sont en train de le faire ou le prévoient à plus long terme.

Mais qu'en est-il véritablement dans le cas de Vince ?

## **Matt Cutts affirme que Vince n'a rien à voir avec les marques**

La question de l'internaute à Matt Cutts était explicite : Pouvez-vous vérifier si Google accorde plus de poids aux marques dans les classements ? Sur ce point, la réponse de Matt est assez claire :

*"Chez Google, au moins au sein de l'équipe en charge de l'algorithme, nous ne pensons pas réellement en termes de "marques". Nous réfléchissons sur les concepts de "confiance", "autorité", "reputation", "Pagerank", "Haute qualité".*

*La philosophie de Google sur les résultats de recherche est restée la même, pratiquement depuis toujours : si quelqu'un vient sur la page de Google et entre "X", nous voulons renvoyer des informations de grande qualité sur X. Et parfois c'est une recherche sur une marque, parfois c'est une recherche d'informations, parfois c'est une recherche de navigation, parfois c'est une recherche transactionnelle.[...]*

*Je ne vois pas cela comme donner plus de poids aux marques. Nous ne pensons pas les "marques" comme si utile que ça en matière de qualité des moteurs de recherche. Par exemple si vous tapez "eclipse", si Google portait vraiment son attention sur les marques nous classerions "Mitsubishi Eclipse" en première position, ou quelque chose de ce genre. Mais si vous allez vraiment taper la requête "eclipse" sur Google, on voit "eclipse.org" en première position, parce que c'est un environnement de développement. On voit aussi le site de la Nasa sur les éclipses, et ensuite des résultats commerciaux. Par exemple, "eclipse" est le nom de l'un des livres de la série "crépuscule", c'est pour cela que nous voyons apparaître une page du site Amazon. Non, nous n'essayons pas de toujours renvoyer des marques. Nous essayons de renvoyer tout ce que nous pensons constituer les meilleurs résultats pour les utilisateurs."*

## **Une étude sur l'hypothèse d'un critère favorisant les marques**

Nous nous sommes livré à une petite étude sur les pages de résultat de Google pour déterminer :

- Si des sites de marques avaient véritablement été favorisés par Vince.
- Et si l'on observait des changements sur des requêtes de marque sur la période "suspecte" (courant janvier 2009).

Nous avons ainsi étudié en particulier l'index de Google.fr pour voir si Vince avait impacté les SERPs telles que nous les voyons en France.

### **Les sites de marques ont ils plus progressé que les autres sur la période ??**

Pour voir si des sites de marque ont véritablement été affectés, nous avons étudié les données issues de la base Marketstory® d'Aposition. Voici les évolutions observées entre début janvier et fin janvier :

Site (domaine)	Pos en + sur page 1	% d'évolution / pos. totales
www.dailymotion.com	1541	20%
www.evene.fr	719	16%
www.commentcamarche.net	476	2%
www.bourse-des-vols.com	394	107%
www.lemonde.fr	323	19%
www.20minutes.fr	306	73%
www.lepoint.fr	298	58%

www.france-info.com	285	108%
www.romandie.com	273	98%
www.touslesprix.com	270	24%
video.google.com	249	16%
www.7sur7.be	214	77%
www.lefigaro.fr	209	10%
www.wat.tv	192	8%
www.webrankinfo.com	188	9%
www.marche.fr	179	12%
www.prixmoinscher.com	171	24%
www.google.com	170	12%
forums.futura-sciences.com	162	6%

Le tableau montre les vingt domaines qui ont le plus progressé en termes de positions obtenues sur la première page de résultats de google entre la semaine 2 et la semaine 4 de 2009. La colonne de droite indique l'évolution en pourcentage du nombre de positions déjà présentes en première page. Cette étude est effectuée sur un échantillon de 200 000 requêtes choisies parmi les plus tapées par les internautes.

Dans l'absolu, on voit ressortir des sites importants, très connus, voire des sites de marque. Mais une comparaison avec le "top des progressions" sur d'autres périodes montre que ce type de sites est régulièrement en tête des progressions. C'est encore plus net si on classe les sites par pourcentage de progression, on s'aperçoit que la liste ne contient plus de sites que l'on peut rattacher à des marques importantes connues :

Site (domaine)	Pos en +	% d'évolution / sur page 1 pos. Acquisées
communiqués.vnunet.fr	58	2900%
www.jobisjob.fr	84	2100%
www.vstars.tv	121	1728%
guide-auto.aol.fr	43	1433%
www.annuaire-chambre-metier.fr	44	1100%
www.live365.com	50	1000%
www.auxpetitesmains.net	39	650%
www.footanglais365.com	46	460%
lapresseaffaires.cyberpresse.ca	69	313%
www.clipser.com	46	270%
video.lematin.ch	55	229%
www.fnacspectacles.com	120	226%
www.lematin.ch	149	222%
www.todoojeux.com	42	190%
www.catholique-nanterre.ccf.fr	86	186%
www.mysacamain.com	55	177%
www.itrmanager.com	79	168%
bourse.lerevenu.com	65	162%
www.lesaffaires.com	51	145%

La réponse est clairement "non", les sites ne marque ne progressent pas plus que d'habitude... La même étude réalisée à des moments différents en 2008 montre une variabilité identique... Vince ne produit pas plus de "révolution" sur les positions des sites que la plupart des changements qui interviennent habituellement dans les classements de Google (qu'ils

proviennent de changements dans l'algorithme ou à l'apparition de nouvelles pages ou d'évolutions des sites existants).

### **Etude sur des requêtes de marque**

Maintenant, que se passe-t'il si on étudie les requêtes sur les 100 marques les plus puissantes en ajoutant les requêtes les plus tapées contenant la marque (exemple "basket Nike tn", ce qui donne un échantillon de 514 requêtes) ?

Une étude sur les variations au sein de la première page de résultat sur ces 514 requêtes de marque donne les résultats suivants :

<b>Périodes comparées</b>	<b>Proportion de nouvelles urls en page 1</b>
Sem 50 2008 / Sem 2 2009	22,63%
Sem 2 2008 / Sem 5 2009	20,95%
Sem 5 2008 / Sem 9 2009	19,56%

Entre 19 et 22% d'urls nouvelles apparaissant sur la première page de résultats d'un mois sur l'autre, cela peut sembler beaucoup, mais c'est en fait le rythme habituel des changements sur Google. Il faut souligner que l'on parle d'urls, et non de domaines (beaucoup de domaines conservent une position en première page d'un mois sur l'autre mais avec des changements dans l'url de la page positionnée).

**CONCLUSION** : non, rien n'a changé de manière brutale ces dernières semaines dans les classements sur les requêtes de marque étudiées.

### ***Vince nous apprend une fois de plus que Google gère différemment des catégories entières de requêtes***

Depuis plusieurs années, les référenceurs soupçonnaient Google d'avoir introduit des règles différentes de classement différentes en fonction de la "nature" des requêtes. Un article du New York Times (*Google Keeps Tweaking Its Search Engine* : <http://www.nytimes.com/2007/06/03/business/yourmoney/03google.html?pagewanted=3>) datant de juin 2007 avait révélé par exemple l'existence d'une catégorie de requêtes baptisées "QDF" (*Query Deserves Freshness*) traitées différemment par Google (sur ces requêtes, Google cherche à renvoyer avant tout des pages récentes et des résultats d'actualité). Certains ont vu également dans cet article une approche "QDD" (*Query Deserves Diversity*), mais il me semble que l'approche décrite peut fort bien être "générale" et s'applique à toutes les catégories de requête.

Ici, à propos de la modification Vince, Matt Cutts nous restitue une classification bien connue des spécialistes des outils de recherche (On retrouve la même classification dans l'ouvrage le plus connu (Chap 19) : *Introduction to Information Retrieval*, Cambridge University Press. 2008 ; Ce qui est flagrant, c'est que Matt Cutts nous les restitue dans l'ordre habituel...) :

- les requêtes de recherche d'information : l'utilisateur cherche des informations générales sur un sujet (exemple : Provence) ;
- les requêtes de navigation : l'utilisateur cherche à trouver une page précise et se sert du moteur pour "naviguer" sur le web ;
- les requêtes transactionnelles : l'utilisateur cherche à faire quelque chose (acheter, télécharger, accéder à un service) ;
- Il existe aussi des requêtes inclassables, situées dans des "zones grises", que l'on ne peut pas catégoriser sans erreur potentielle sur l'intention de l'utilisateur.

Vince serait donc un cas avéré de plus d'un traitement particulier pour une catégorie de requêtes. Compte tenu des requêtes identifiées comme affectées par Vince, on peut clairement identifier qu'il s'agit de requêtes de type "transactionnelles", même si on ne peut

pas exclure que la classification de google soit plus subtile et se base sur une logique de classification différente, ce qui signifie que certaines requêtes de navigation ou de recherche d'information puissent être affectées.

Ce qui amène à une révolution dans l'approche classique du référencement, qui consistait à optimiser les pages sur deux familles de critères : des critères dépendants de la requête, et des critères indépendants de la requête.

Aujourd'hui il s'agit plutôt de raisonner selon trois familles de critères, la deuxième se manifestant tel Vishnu par toute série d'avatars assez différents.

- Les critères liés à la requête (exemple la similarité entre le texte de la requête et les documents) ;
- Les critères liés à la catégorie de requête (exemple la date de publication d'un article ou d'un billet sur les requêtes QDF) ;
- Les critères 100% indépendants de la requête (comme le PageRank, du moins dans sa version "classique").

### ***Sur quoi se base Vince pour faire monter certains sites ?***

Là, nous en sommes réduits à faire des hypothèses. Le simple fait de tourner légèrement un bouton vers la droite pour augmenter le poids de certains critères peuvent aboutir aux résultats observés. Mais il est difficile d'exclure aussi l'intégration de nouveaux critères dans l'algorithme. Quelqu'un nous a rappelé à ce sujet l'étrange "bug" survenu le jour d'Halloween 2008 : pendant quelques heures les SERPs ont présenté des résultats étranges, avec par exemple un wikipedia rétrogradé parfois en deuxième page sur certaines requêtes. Bug ou tentative ratée d'introduire un nouveau critère "de qualité" ?

Quels sont les critères qui peuvent, en tout cas, produire ces résultats ?

- Un coup de pouce sur certains critères révélant le "sérieux" du site (age du domaine, historique, ...) ;
- Un coup de pouce sur le Trustrank ;
- Une utilisation renforcée des *Topic Sensitive Pageranks* (mesure de "topicality") ;
- L'introduction de critères d'usage : trafic estimé, taux de rebond observé etc. (Matt Cutts vient encore de réaffirmer le 2 avril 2009 au PubCon d'Austin Texas que ces indicateurs généraient du bruit et n'étaient pas utilisés. Mais certains brevets déposés par Google démontrent qu'ils peuvent être utilisés d'une façon efficace (mais différente de celle que le sens commun nous conduit à envisager)).
- L'utilisation d'indicateurs issus des médias sociaux (mais certains des sites "boostés" n'ont guère de chance de voir régulièrement la communauté Twitter buzzer à leur propos, c'est donc sujet à caution).
- Un indicateur de réputation : une mesure du nombre de fois ou l'on parle positivement du site ou de la page
- ...

Certains ont évoqué un travail sur les entités nommées. Dans ce cas, il s'agit forcément d'un travail sur des relations entre une entité nommée et son environnement sémantique naturel. C'est sans doute quelque chose sur lequel Google pourrait travailler, mais cela produirait d'autres résultats.

### ***Faut-il se préoccuper de cette modification de l'algorithme ?***

Compte tenu de son impact limité, voire insensible dans le cas de sites positionnés sur un grand nombre de requêtes, il ne paraît pas utile d'engager un chantier particulier en réaction à cette modification de l'algorithme. Qui plus est, il semble qu'elle ait pour effet d'améliorer effectivement la pertinence des résultats perçus par l'internaute. Ce ne sera sans doute pas l'avis des webmasters qui auront été évincés de leurs positions favorites...

Mais on assiste apparemment à un renforcement des positions de sites performant correctement sur les critères de confiance, d'autorité, de réputation, de qualité. Il s'agit d'une

tendance lourde de l'évolution de l'algorithme dont il faut tenir compte dans une stratégie à long terme.

Cette "amélioration Vince" est malgré tout anecdotique. Force est de constater qu'il existe encore de très nombreuses requêtes où les pages de résultats de Google ressemblent encore à une "fosse d'aisance" pour reprendre l'image d'Eric Schmidt, remplie de pages de spam ou de pages de qualité médiocres.

***Philippe Yonnet, Directeur Technique @Position (<http://www.aposition.com>) et président de l'association SEO Camp (<http://www.seo-camp.org/>)***

Réagissez à cet article sur le blog des abonnés d'Abondance :  
<http://abonnes.abondance.com/blogpro/2009/04/vince-une-mise-jour-reellement-majeure.html>