

Vince : Une mise à jour de Google qui reste un mystère, quatre ans après

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

La mise à jour Vince de Google, qui a été détectée en 2009, a toujours été un mystère pour le monde du SEO. Après avoir imaginé qu'elle mettait en avant les sites de marques reconnues, d'autres hypothèses ont été émises. Qu'en est-il aujourd'hui et sait-on réellement, quatre ans plus tard, ce que le moteur de recherche a mis en place avec ce Vince qui a fait couler beaucoup d'encre ?...

La mise à jour Google Vince (2009) fait partie des changements qui ont été les plus commentés (et vilipendés) par la communauté SEO ces dernières années. Beaucoup y ont vu un "tournant" dans l'attitude de Google vis à vis des sites web indexés, en démontrant la volonté du moteur de favoriser les sites de marque.

Mais qu'en est-il vraiment ? Avec le recul, sait-on mieux ce qui s'est passé en 2009 dans les pages de résultats ?

Nous allons voir que les explications jugées les plus plausibles au début sont probablement totalement fausses, et que finalement, on sait toujours peu de choses sur Vince. Un exemple qui doit nous laisser prudents sur les capacités de la communauté SEO à décoder ce qui se passe chez Google...

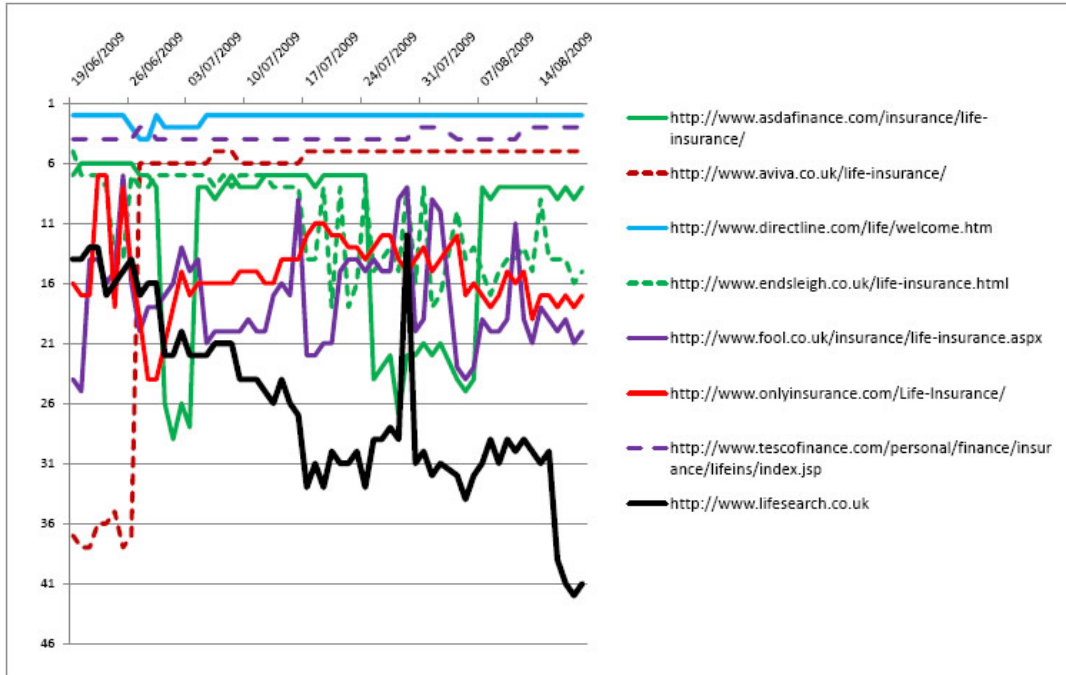
Petit rappel historique

La mise à jour Vince a été détectée pour la première fois grâce à un message daté du 20 février 2009 et publié sur Webmasterworld, le forum américain sur le webmastering. 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... Le phénomène a ensuite été confirmé par plusieurs gourous du SEO, au premier rang desquels on trouve Aaron Wall de SeoBook. Il semble qu'en fait, le changement ait eu lieu un mois avant, soit le 18 janvier 2009.

Matt Cutts, le responsable de l'équipe "search quality" chez Google, a confirmé qu'il y avait bien eu un changement dans l'algorithme, dans une vidéo réalisée le 26 février 2009 et diffusée sur Youtube (*voir URL à la fin de l'article*). 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. Cette mise à jour était donc ainsi baptisée...

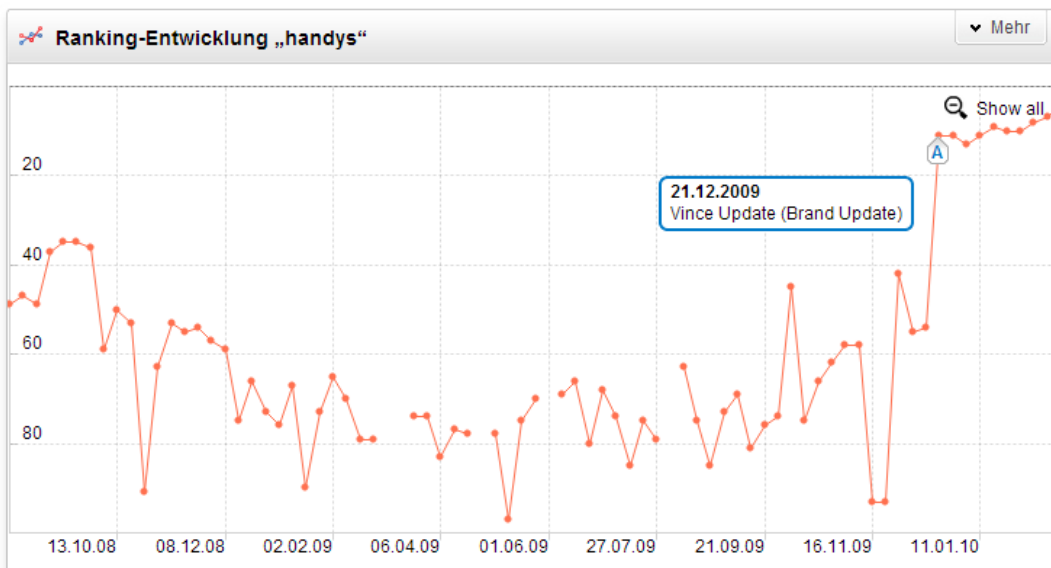
Un déploiement progressif dans d'autres pays

Si la mise à jour date de janvier 2009 aux USA, il a fallu attendre plusieurs mois pour qu'elle soit étendue aux autres pays de langue anglaise. Au Royaume Uni, le changement est détecté en juillet 2009 sur Google.co.uk . Il semble que le changement date du 26 juin 2009 en réalité.



Graphes de positions réalisés par l'agence anglaise econsultancy pour mettre en évidence l'impact de Vince sur la requête : « Life Insurance »

En France, la date d'arrivée de la mise à jour n'a pas été réellement mise en évidence. Elle semble remonter à la fin de l'année 2009 ou au début de l'année 2010. En Allemagne, grâce à une étude de la société Sistrix, on sait que Vince a été probablement déployé fin décembre 2009.



L'un des exemples fournis par Sistrix pour montrer que la date du changement Vince correspond pour l'index allemand au 21 décembre.

Keyword	Neue Markendomains
jobs	arbeitsagentur.de (Neu)
computer	alternate.de (Neu), acer.de (+10)
handys	t-mobile.de (Neu), vodafone.de (+30), nokia.de (Neu), sonyericsson.com (+50)
baby	pampers.de (+34)
drucker	canon.de (Neu), brother.de (Neu), lexmark.de (+10)
dsl	alice-dsl.de (+10), 1und1.de (+16)

Les sites de marque qui ont fait une remontée notable dans les pages de résultats en Allemagne le 21 décembre 2009 (source Sistris)

Vince favorise-t-il réellement les sites de marque ?

Ce qui a tout de suite frappé les observateurs avec Vince, c'est la remontée brutale dans les pages de résultats de sites de marques ou de compagnies ayant « pignon sur rue », au détriment d'autres sites.

Voici un exemple des changements caractéristiques constatés sur la requête "airline tickets" après Vince. 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 violemment progressé à la même date, mais n'est entrée sur la première page que le 1er février.

- #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

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 (travelocity).

Un tel phénomène a été remarqué sur de nombreuses requêtes à l'époque. C'est pourquoi Vince a été surnommée "the brand update", et même "the big brand update", la mise à jour "grandes marques".

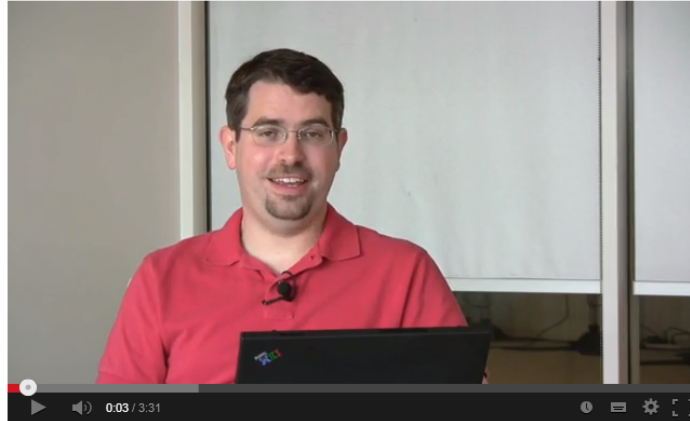
En fait, Vince n'avait rien à voir avec les marques

Nous avons pu réaliser une étude en avril 2009 sur un échantillon de requêtes sur les marques les plus connues (requêtes du type "basket Nike"). Cette étude avait révélé que Vince n'avait pas chamboulé les pages de résultats sur ces mots clés. Même chose sur les requêtes "appelant des réponses de sites de marque" (requêtes du type « vols secs »). Cette étude avait été partiellement publiée dans la lettre professionnelle d'Abondance. Matt Cutts a l'époque avait lâché que le "changement Vince" (Matt Cutts semble ne pas aimer le terme "mise à jour" si l'infrastructure n'est pas réellement modifiée) n'affectait

en réalité qu'un faible nombre de requêtes. Ce faible impact empêchait donc les analyses statistiques de produire des résultats probants.

Et Matt Cutts avait bien confirmé (dans sa vidéo du 26 février 2009) que la notion de "marque" n'avait pas grand sens dans le contexte de Google :

"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", "réputation", "Pagerank", "Haute qualité".



Is Google putting more weight on brands in rankings?

Une capture d'écran de la vidéo du 26 février 2009 (datée du 4 mars) où Matt Cutts répond aux allégations d'Aaron Wall de SEO Book

Matt Cutts a eu l'occasion de reconfirmer de manière encore plus claire sa position dans un tweet publié en novembre 2010 :



@aaronwall seriously though, the Vince update was not about brands. Believe me or not, but that's so.

Vince et le reranking sur les critères de "confiance"

Début 2009, les informations lâchées par des ingénieurs de Google et par Matt Cutts laissaient penser que le "changement" avait quelque chose à voir avec la notion de "trust" (confiance). Vince aurait donc été le reflet d'un poids accru des indicateurs de confiance et d'autorité dans l'algorithme.

Un brevet publié par Google fin 2009 est venu confirmer l'intérêt de Google pour ces notes de confiance (voir référence dans la Bibliographie). Ce brevet décrit une nouvelle notion de « Trustrank », qui n'a rien à voir avec la notion de Trustrank dont parlent habituellement les référenceurs (ce Trustrank-là, rappelons-le, est en fait un concept qui a été décrit dans un brevet publié par ... Yahoo ! et non Google).

Début 2009, la communauté SEO évoque donc majoritairement cette explication comme la bonne, et on voit fleurir les recommandations du type : « pour mieux se positionner depuis Vince, il est important de travailler ses backlinks pour acquérir plus de confiance et d'autorité ».

Vince et l'exploitation des logs de requête pour améliorer les résultats

En juillet 2009, le déploiement de Vince au Royaume Uni amène nos confrères britanniques à observer à leurs tours les changements provoqués par Vince sur certaines requêtes. Chez Icrossing, Econsultancy, Distilled, les consultants s'aperçoivent que la date du changement est difficile à isoler au premier coup d'œil. La plupart s'accordent sur la date du 26 juin 2009, mais constatent que les positions continuent à bouger depuis...

Et voilà qu'un ingénieur de Google, Matthew Trehella, lâche une bombe au cours de l'expo A4U le 13 octobre 2009 : Vince serait en réalité lié à la mise en place par Google d'un système cherchant à exploiter le comportement de recherche des utilisateurs pour réordonner les résultats.

Matthew Trehella aurait déclaré : « *chaque fois qu'un utilisateur doit effectuer une deuxième recherche, Google considère cela comme un échec pour ne pas avoir su procurer le résultat correct la première fois* ».

En clair, Vince ne serait qu'une application de l'exploitation des logs de requête et du tracking des clics pour améliorer les résultats. Une méthode bien connue par tous les spécialistes de moteurs de recherche...

Rappel du principe de l'analyse des sessions de recherche

L'analyse des logs de requête et des actions (clics) des utilisateurs a été décrite depuis longtemps (fin des années 90) comme un moyen théorique d'améliorer la pertinence des résultats de manière efficace.

Il est en effet possible d'enregistrer dans un fichier (log) les requêtes et les actions successives effectuées par un internaute au cours de la même session de recherche. Une « session de recherche » étant une série de requête et d'actions réalisées depuis la même IP et séparées par intervalle de temps réduit.

Voici un exemple tiré d'un fichier log du moteur Excite (a présent disparu) qui permet de bien comprendre ce que l'on conserve comme information et quelles conclusions on peut en tirer d'un point de vue « retour de pertinence » :

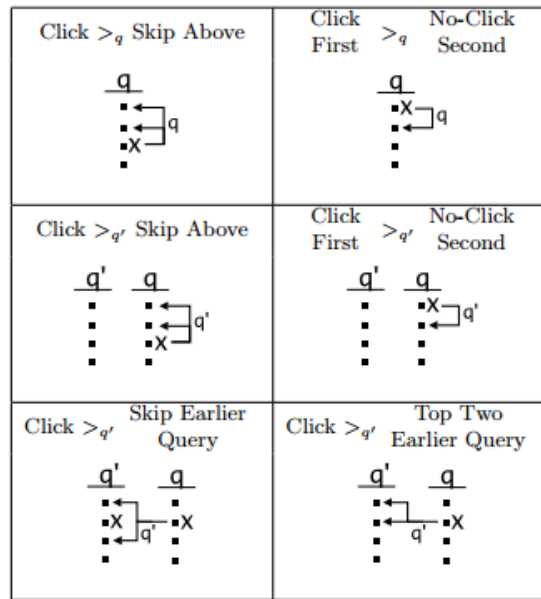
[10/09 06:39:25] Query: holiday decorations [1-10] ->consultation page une : pas d'action
[10/09 06:39:35] Query: [web]holiday decorations [11-20] ->page 2, pas d'action
[10/09 06:39:54] Query: [web]holiday decorations [21-30] ->page 3
[10/09 06:39:59] Click: [webresult][q=holiday decorations][21] -> clic sur un résultat <http://www.stretcher.com/stories/99/991129b.cfm>
[10/09 06:40:45] Query: [web]halloween decorations [1-10] ->Deuxième requête : la première requête n'a pas permis de trouver la page, ou l'information, une deuxième requête est nécessaire...
[10/09 06:41:17] Query: [web]home made halloween decorations [1-10]
[10/09 06:41:31] Click: [webresult][q=home made halloween decorations][6] http://www.rats2u.com/halloween/halloween_crafts.htm -> deuxième clic
[10/09 06:52:18] Click: [webresult][q=home made halloween decorations][8] <http://www.rpmwebworx.com/halloweenhouse/index.html> -> troisième clic
[10/09 06:53:01] Query: [web]home made halloween decorations [11-20] ->page 2
[10/09 06:53:30] Click: [webresult][q=home made halloween decorations][20] <http://www.halloween-magazine.com/> ->fin de la session, abandon ou info trouvée !

L'analyse combinée des requêtes tapées par les internautes et des liens cliqués dans les pages de résultats permet de tester en vraie grandeur la qualité des résultats proposés. En effet, si sur une requête donnée, un nombre important d'utilisateurs clique sur le résultat situé en 7e position, c'est donc qu'ils jugent ce résultat plus pertinent. Lorsque ce

phénomène se produit sur un nombre élevé de requêtes du même type, c'est le signe que l'algorithme de classement du moteur est pris en défaut, et qu'il faut l'améliorer.

L'exploitation des logs de requête et des sessions de recherche est un domaine qui a beaucoup progressé depuis 2009 et qui a aujourd'hui des applications multiples et variées dans les moteurs de recherche.

Elle a été « théorisée » par Thorsten Joachims de l'université de Cornell dans les années 2004 à 2005. Le Pr Joachims a notamment expliqué comment effectuer les calculs nécessaires pour réaliser un système performant pour de gros volumes de données (comme ceux rencontrés chez Google ou Bing).



L'illustration ci-dessus montre les décisions prises par un système de réordonnement des résultats s'appuyant sur les sessions de recherche :

1. Si un résultat est cliqué, alors ce résultat est jugé plus pertinent sur la requête que les résultats au-dessus ;
 2. Si un résultat est cliqué en position 1, alors il est jugé plus pertinent sur la requête que le résultat en position 2
 3. Si un résultat est cliqué sur la page 2 ou une requête raffinée, alors ce résultat est jugé meilleur que tous les liens de la page situé au-dessus...
- Etc.

Vince est-il un effet d'un système d'amélioration des résultats basé sur les sessions de recherche ?

Cette explication est plausible : elle rend compte du changement progressif des positions sur les requêtes impactées par Vince longtemps après le changement. Et elle explique la remontée de résultats de sites très connus également : on peut imaginer qu'un tel système favorise les sites disposant d'une forte notoriété spontanée.

Peut-on pourtant être sûr qu'il s'agisse de la « vraie » explication derrière le changement Vince ? Pas du tout !

Les propos de Matthew Trehella n'ont d'ailleurs pas confirmés : ils ont été rapportés par Patrick Atloft, un gourou du SEO au UK, sans doute digne de confiance, mais il s'agit au mieux d'une interprétation de commentaires fait par l'ingénieur de Google. Pas un « verbatim » donc. Par ailleurs, Matthew Trehella n'était probablement pas autorisé à lâcher cette explication, et rien n'indique qu'il fasse partie des équipes qui sont réellement

dans la confiance à propos des changements algorithmiques de Google. En réalité, il est « developer advocate » chez Google UK.

Par ailleurs, il ne fait aucun doute aujourd'hui que Google exploite les sessions de recherche pour améliorer ses pages de résultats. Quand a-t-il commencé à le faire ? Etait-ce début 2009 avec Vince ? Ou avant ? Car après tout, cette idée a été théorisée à une époque contemporaine du lancement de Google, et les moyens de « calculer » en temps quasi réel les indicateurs permettant de « reranker » les résultats en fonction des clics des internautes existaient depuis au moins 5 ans en 2009 ! Dans ce cas, Vince serait-il un simple changement apporté à ce système ? Pourquoi pas...

Quatre ans après : des pistes, mais pas de certitudes

Finalement, avec le recul, on ne sait pas grand-chose avec certitude sur cette mise à jour. On sait que Vince ne cherchait pas à favoriser des marques, mais juste à améliorer la pertinence sur certaines requêtes. La piste d'un boost d'un indicateur de confiance ou d'autorité a longtemps semblé être la bonne, mais elle est sans doute trop simpliste. La piste du reranking en fonction des clics utilisateurs semble plus fondée, mais en réalité, elle n'a jamais été confirmée (ni infirmée d'ailleurs).

Ce retour en arrière nous permet de nous interroger aussi sur les limites de l'exégèse des propos de Matt Cutts (responsable Search Quality, en charge de la lutte contre le spam) ou d'Amit Singhal (responsable Qualité de Recherche, en charge de la qualité de l'algorithme). On peut s'interroger aussi sur la manière peu scientifique avec laquelle les mises à jour sont analysées, et sur la sensibilité à la rumeur de la communauté SEO (incontestablement très forte : l'absence d'infos officielles crée un boulevard pour les fausses informations).

Un constat qui doit nous amener à faire preuve de la plus grande prudence à la lecture des analyses sur les dernières mises à jour de Google : Hummingbird et Penguin 2.1 !

Bibliographie

Brevets

Search result ranking based on trust

<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fmetahtml%2FPTO%2Fsearch-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7,603,350.PN.&OS=pn/7,603,350&RS=PN/7,603,350>

Invented by Ramanathan Guha : Assigned to Google, US Patent 7,603,350, Granted October 13, 2009, Filed: May 9, 2006

Vidéos

Matt Cutts : Is Google putting more weight on brands in rankings?

<http://youtu.be/LMfWPWUh5uU>

Articles

Brands vs Query Refinement: Is Google Using The Second Search?

<http://www.seobook.com/brands-vs-query-refinement-google-using-second-search>

The brand update is about maximising satisfaction rates

<http://www.branded3.com/blogs/the-brand-update-is-about-maximising-satisfaction-rates/>

Unlocking Google's vince update

http://connect.icrossing.co.uk/unlocking-googles-vince-update_2990

Vince unravelled : does Google recognise brand equity ?

<http://econsultancy.com/fr/blog/4862-vince-unravelled-does-google-recognise-brand-equity-part-1>

Sistrix : Google Vince Update

<http://www.sistrix.de/frag-sistrix/google-algorithmus-aenderungen/google-vince-update/>

Publications scientifiques

Search Engines that Learn from Implicit Feedback : Joachims, Radlinski

<http://luci.ics.uci.edu/websiteContent/weAreLuci/biographies/faculty/djp3/LocalCopy/04292009.pdf>

Optimizing Search Engines Using Clickthrough Data : Joachims

http://pdf.aminer.org/000/472/865/optimizing_search_engines_using_clickthrough_data.pdf

Philippe YONNET , *Directeur de l'agence Search-Foresight / Groupe MyMedia.*
Président de l'association SEO Camp (<http://www.seo-camp.org/>)