

10 Questions à la "Search Quality Team" de Google

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

10 questions posées à 5 personnes de l'"Equipe qualité de recherche" de Google (Matt Cutts, John Mueller, Susan Moskwa, Natalie Peret, Anouar Bedahou) : tel est le but de cette interview qui traite de nombreux sujets importants dans le cadre du référencement naturel d'un site web : indexation et syntaxe "site:", présence de mots clés dans les URL, temps de chargement des pages et temps de réaction d'un serveur, analyse et compréhension des codes JavaScript, mise à jour Mayday du moteur, balise link rel="canonical", prestations de conseil, disparition hypothétique du PageRank, perte éventuelle de PageRank dans les redirections 301 et avenir du positionnement, autant de points intéressants sur lesquels Google donne ici son opinion...

La "Search Quality Team" ou "Equipe qualité de recherche" de Google communique de plus en plus avec les webmasters, que ce soit lors de "site clinics" dans des salons, ou en intervenant sur des forums spécialisés, dans des interviews ou, par exemple, dans les "questions/réponses" sur le site Abondance (voir page d'accueil à l'adresse <http://www.abondance.com/>). Les réponses qui sont fournies aux questions posées dans cette interview sont souvent intéressantes et argumentées, même si elles amènent parfois quelques commentaires, que nous nous sommes permis d'ajouter après chacune d'elles...

Nous avons donc proposé à cette "Equipe qualité de recherche" de répondre à 10 questions qui nous semblaient très importantes au sujet du référencement de sites web. Vous les trouverez ci-dessous, ainsi que les réponses de cinq personnes de cette "spam brigade", à savoir :



Matt Cutts
Head of Google
Webspam team
(Mountain View,
US et parfois
Kilimandjaro :))



John Mueller
Webmaster Trends
Analyst
(Zurich, Suisse)



Susan Moskwa
Webmaster
Trends Analyst
(Mountain View,
US)



Nathalie Peret
Search Quality
Strategist
(Dublin, Irlande)



Anouar Bendahou
Search Quality
Associate
(Dublin,
Irlande)

Sachez que l'interview s'est effectuée par mail, en anglais et que les réponses ont été traduites par nos soins (cette traduction a cependant été transmise à Google et approuvée par ses équipes). Une version originale (en anglais) des questions/réponses [est disponible ici](http://abonnes.abondance.com/archives/2010-09/interview-google-en.doc) (<http://abonnes.abondance.com/archives/2010-09/interview-google-en.doc>).

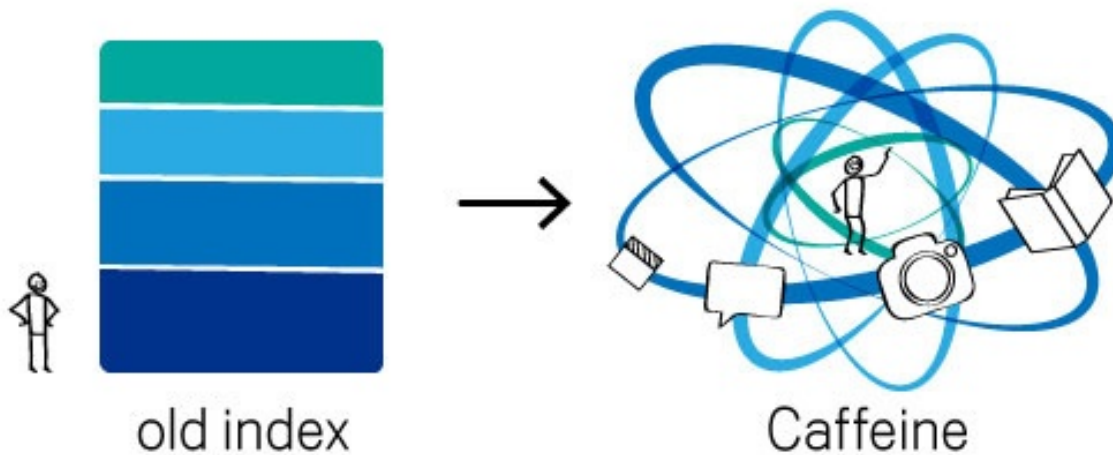
- De nombreux problèmes sont souvent constatés par les webmasters avec la syntaxe "site:" de Google, sensée fournir le nombre de pages d'un site web référencées par le moteur. Ce type de requête (par exemple : *site:www.abondance.com*) génère le plus souvent un nombre de résultats très différent d'un jour à l'autre et semble notoirement erroné. Il est ainsi extrêmement difficile de suivre le nombre de pages indexées par Google pour un site web donné. On sait que, d'autre part, la syntaxe "link:" de Google est, de la même façon erronée et ne donne qu'un nombre de backlinks très limité sur le moteur, mais le problème est contourné par des données beaucoup plus fiables sur ces backlinks dans les Webmaster Tools. Dans ce cas, ne serait-il pas possible d'avoir, dans ces mêmes Webmaster Tools, des informations plus fiables et exhaustives sur l'indexation des pages pour pallier le manque de fiabilité de l'opérateur "site:" ?

Les webmasters nous posent en effet souvent cette question, notamment dans notre forum d'aide pour les webmasters (<http://www.google.com/support/forum/p/webmasters?hl=fr>) pour nous demander s'il ne s'agissait pas là d'un bug. En fait, l'opérateur "site:" a été prévu au départ pour fournir un nombre approximatif, une estimation des pages d'un site indexées par Google. Ce chiffre est intéressant pour vérifier la "santé" de votre site mais il est clair qu'il ne donnera pas un chiffre précis, une vision exacte de l'indexation de vos pages. Des fluctuations sont à attendre et les chiffres renvoyés vont effectivement changer selon les jours.

Sachez également que le "switch" entre notre ancien index et Caffeine a généré certaines différences entre les recherches/requêtes avec l'opérateur site :...

Pour obtenir des résultats plus précis et plus fiables, nous ne pouvons que vous conseiller de regarder plutôt l'indicateur du nombre d'URL de votre Sitemap indexées, dans les Webmaster Tools. Cet outil vous donnera également d'autres indications sur les erreurs de crawl (<http://www.google.com/support/webmasters/bin/answer.py?answer=35120>) que nos robots ont pu rencontrer sur votre site. Enfin, selon nous, le nombre de pages indexées nous semble un indicateur moins intéressant que le positionnement de ces pages, leur taux de clic ou leur taux de conversion.

Regarder et fournir des indications dans les Outils pour les webmasters sur les pages indexées (et non incluses dans un Sitemap) est un point qui a été discuté plusieurs fois dans notre équipe. Pour l'instant, ce n'est pas une priorité mais c'est quelque chose que nous regardons et que nous n'avons pas abandonné pour autant...



L'architecture Caffeine vue par Google (<http://googleblog.blogspot.com/2010/06/our-new-search-index-caffeine.html>)

Notre opinion : le nombre de pages indexées ne serait pas un bon indicateur pour le référencement d'un site web ? Objection, votre honneur :-)... Si, bien évidemment, il ne peut pas être le seul, il reste selon nous un critère important qui permet plusieurs choses :

- se rendre compte du pourcentage de page indexées par rapport au nombre total de pages ;
- détecter des soucis techniques (URL non compatibles, etc.) sur certaines pages ;
- etc.

D'autre part, il est évident que le taux de clics, le taux de conversion et le positionnement sont des indices importants, mais sur une page non indexée, difficile d'en voir l'intérêt... L'indexation reste, qu'on le veuille ou non, l'étape "zéro" du référencement et il nous semble important de le mesurer de la façon la plus fiable possible.

Google donne, dans réponse, une façon d'obtenir des résultats "plus fiables" en relevant le nombre d'URL indexées par rapport au nombre d'URL soumises dans le Sitemap, chiffre fourni dans les Webmaster Tools. Or, l'expérience nous pousse à penser que ce chiffre-là est également notoirement faux, et un phénomène fréquent montre un nombre d'URL indexées équivalent à environ la moitié du nombre d'URL soumises dans le Sitemap. Pourquoi ? Mystère...

Bref, peut-être faudra-t-il faire une pétition pour avoir droit un jour à ces chiffres fiables d'indexation de site web dans les Webmaster Tools ? Ou cela pose-t-il des problèmes techniques insolubles à Google de fournir de telles données ? En complément, voici un post sur le forum d'aide Google pour les webmasters qui parle de ce sujet :

<https://sites.google.com/site/webmasterhelpforum/fr/aide-au-referencement/interpreter-les-donnees-de-site-et-de-sitemap>

- De nombreux webmasters utilisent la réécriture d'URL pour insérer certains mots clés importants dans l'adresse de leurs pages. Mais certains doutent également de l'importance et de l'utilité de telles actions dans le cadre d'une stratégie SEO. Quel est votre point de vue à ce sujet ? La présence de mots clés dans les URL est-elle un point essentiel pour le positionnement d'une page web dans les Serps de Google ?

La présence de certains mots pertinents dans l'URL peut légèrement aider, mais pas au point de "bourrer" les URL de tonnes de mots clés... Des URL comportant trop de requêtes potentielles peuvent sembler "spammy" aux utilisateurs qui risquent de ne pas cliquer par manque de confiance. Quelques termes très pertinents seront bien plus efficaces pour le SEO et aideront même l'internaute.

Aussi, si vous créez un site web, faites en sorte que vos URL contiennent dès le départ de 1 à 3 mot(s) clé(s) pertinent(s) dans chaque URL et ce sera parfait. Une réécriture d'URL de ce type par votre CMS donnera d'excellents résultats.

En revanche, si votre site existe déjà et que ses URL ne sont pas réécrites (par exemple si elles contiennent des paramètres), nous conseillons de les laisser telles quelles ou de faire en sorte que d'éventuels changements ne posent pas de problèmes de crawl ou d'indexation à nos robots pour cause de réécritures mal mises en place.

Nous recommandons fortement de lire notre "Guide d'optimisation pour les moteurs de recherche" (<http://www.google.com/intl/fr/webmasters/docs/search-engine-optimization-starter-guide-fr.pdf>) à ce sujet, car il donne plus de détails sur la façon d'optimiser votre site, et notamment la structure des URL pour qu'elles soient facilement explorables par les moteurs de recherche...

Notre opinion : la réponse de Google semble en effet correspondre à ce que nous constatons au quotidien : des URL réécrites "proprement", sans excès, peuvent donner de bons résultats. Dès que le "bon sens", si important en référencement :), est bousculé, les résultats sont moins évidents et le "spamming" (et donc la pénalité) n'est pas très loin...

Attention également à la réécriture erronée qui pose souvent des problèmes, notamment dans la sauvegarde de l'existant : des redirections 301 doivent être mises en place entre les anciens et les nouveaux intitulés d'URL pour que la version préalable de l'URL reste compatible avec son intitulé réécrit... Il sera souvent bon de demander conseil à un expert de l'"URL rewriting" pour s'assurer que tout va bien de ce côté-là, car c'est assurément l'un des points les plus sensibles lors de la mise en place d'une stratégie SEO sur un site déjà en ligne...

- Depuis plusieurs mois, on sait que la vitesse d'un site web, son temps de réaction, le temps d'affichage des pages sont des critères de pertinence pour le référencement d'un site web. Nous savons également que ce n'est encore qu'un critère mineur, touchant à peine 1% des sites actuels. Mais est-ce que cela est pris en compte par Google comme un "malus" pour les sites lents ou comme un "bonus" pour les sites rapides ?

Pour les internautes, les sites qui s'affichent rapidement rendent le surf plus facile et plus agréable. Il peut s'avérer frustrant pour les utilisateurs du réseau d'avoir à attendre le chargement d'une page pour obtenir une information. Dans ce cas, un site rapide sera un "plus" indéniable pour tout le monde et pour Google en particulier. Comme vous le dites, ce critère affecte un faible pourcentage des résultats (1% des requêtes uniquement) et est donc un critère parmi bien d'autres... Néanmoins, un site lent peut également être affecté négativement par notre algorithme. Mais nous étudions la possibilité pour les sites rapides

d'être "récompensés" d'une façon ou d'une autre. Les webmasters peuvent vérifier la performance de leurs pages et de leur site dans les "laboratoires" des Outils pour les webmasters (<http://www.google.com/support/webmasters/bin/answer.py?answer=158541>). Ils y trouveront également des indications et des suggestions pour accélérer et améliorer ce critère. Vous trouverez également d'autres outils pour tester et améliorer la vitesse de chargement de vos pages à cette adresse : <http://googlewebmastercentral.blogspot.com/2010/04/using-site-speed-in-web-search-ranking.html>

Notre opinion : à notre avis, le fait de vérifier la vitesse de réaction de votre serveur et la rapidité d'affichage de vos pages servira bien plus à l'internaute, qui attendra moins que les données que vous lui proposez s'affichent, qu'aux moteurs de recherche, pour lesquels ce critère reste encore très mineur... Mais un site web étant d'abord créé pour les internautes, n'est-ce pas un point important que vos pages s'affichent le plus vite possible ? Et si, parfois, cela peut améliorer votre référencement, c'est aussi bien. Mais cela ne doit pas être votre motivation principale...

- Est-ce que Google comprend et analyse actuellement, mieux que par le passé, le code JavaScript ? Si oui, qu'arrive-t-il à détecter dans ces codes ? Les liens ? Le texte ? D'autres informations ?

Au départ, notre intérêt principal dans l'analyse du JavaScript était d'y découvrir des nouveaux liens, vers des pages que nous ne connaissions pas. Au fil du temps, nous avons fait beaucoup d'efforts pour y découvrir à la fois les liens et les contenus que ces codes contiennent. Nous cherchons chaque jour à être meilleurs dans ces deux domaines. Nous pouvons également détecter d'éventuels attributs *rel="nofollow"* dans le code JavaScript. Cependant, bien que nos efforts soient constants dans ce domaine, nous ne pouvons que vous conseiller de fournir vos contenus et vos liens dans le langage HTML "classique". C'est important pour être sûr qu'aucun problème de crawl et d'indexation n'interviendra par la suite...

Notre opinion : la réponse de Google rejoint l'opinion de nombreux référenceurs, sur le fait que Google interprète de mieux en mieux le JavaScript. Mais il faut également voir que Google a intérêt à faire passer ce message, notamment auprès des référenceurs "black hat" qui utilisent ce langage notamment pour cacher certains liens dans le cadre de stratégie de "Page Rank Sculpting"... Malgré cela, la prudence reste de mise et on peut considérer qu'actuellement, Google a souvent une vision assez claire des contenus insérés dans des codes JavaScript et que, par conséquent, si un webmaster a la volonté de "cacher" un contenu par ce biais, il faudra que la technique utilisée soit assez "costaud". Ce type de pratique risque donc d'être assez rapidement "à risque"...

- La mise à jour "Mayday" de l'algorithme de pertinence du moteur a parfois eu des répercussions importantes sur certains sites. Quelles sont vos suggestions à ce sujet pour les webmasters dont le trafic aurait pâti de cette nouvelle mouture ?

De nombreux sites ont effectivement rapporté une baisse de trafic sur des requêtes de longue traîne après la mise à jour qui a été appelée "Mayday" par certains webmasters. Il faut bien noter que Mayday est un changement algorithmique qui tente de rendre certains sites plus pertinents qu'avant pour les internautes sur certaines requêtes, à la place de certains sites moins intéressants selon nous. Ce filtre est appliqué automatiquement pour améliorer la qualité des résultats de recherche sur les requêtes de type "longue traîne". Si votre site a été affecté, vérifiez les requêtes pour lesquelles vos pages sont moins bien positionnées. A *contrario*, regardez pourquoi certaines autres pages se classent mieux. Contiennent-elles des commentaires ou des critiques utiles pour les internautes ? Ont-elles créé une communauté ? Peuvent-elles être considérées comme "expertes" sur le sujet ? Voici des pistes qui devraient vous aider...

Notre opinion : Selon nous, la mise à jour "Mayday" a été mise en place par Google pour contrecarrer certaines solutions, apparues depuis quelques mois, visant à créer d'innombrables pages reprenant le contenu d'un site, en créant des nuages de mots clés divers et variés, agrégeant des contenus du site pour créer des centaines, voire des milliers de pages web, tentant ainsi d'utiliser le phénomène de "longue traîne" pour obtenir plus de visibilité... Ces pages n'étant que des reprises et du copier/coller de contenus déjà présents dans un site web, elles ne contiennent aucun UGC (*User Generated Content*) - commentaires, avis, critiques, etc. - et ont bien souvent un PageRank très faible. Le but était certainement de baisser l'importance, et donc le ranking, de ce type de page. Idée à confirmer cependant...

- Est-ce que la balise link rel="canonical" est réellement efficace ? On peut s'apercevoir par exemple que, sur Google News, malgré l'utilisation d'une telle balise, la "version imprimable" d'un article est pourtant souvent prise en compte par le moteur au lieu de la version "normale"...

Cette balise est très utile et très efficace pour la recherche "Web". Cependant, Google News fonctionne de façon différente par rapport au moteur Web, aussi cette balise n'est peut-être pas la meilleure façon de désindexer un contenu de notre moteur d'actualité. Pour cela, nous vous recommandons d'utiliser la balise meta robots avec le paramètre "googlebot-news" ou un fichier robots.txt. Pour plus de détails sur ce sujet, n'hésitez pas à lire ces pages : http://www.google.com/support/news_pub/bin/answer.py?answer=93977.

Notre opinion : le fait que Google News ne prenne pas en compte la balise link rel="canonical" est selon nous un bug étonnant et dommageable. En effet, cette balise transfère les liens (backlinks) de la page dupliquée à la page canonique, alors que ce n'est pas du tout le cas lors d'une désindexation au travers de la balise meta "robots" ou du fichier robots.txt. Un problème à résoudre donc par Google...

- Pourquoi Google ne propose-t-il pas des prestations de conseil et d'audit pour certains sites web, notamment de gros sites proposant de nombreuses informations, afin qu'ils soient plus compatibles avec son moteur de recherche ? De nombreux webmasters apprécieraient ce type d'aide (proposée d'ailleurs par Google pour ses liens sponsorisés Adwords). Et ne serait-ce pas un avantage pour vous d'indexer des sites web plus "compatibles" ?

Notre "Search Quality Team" ne propose pas de services de conseil et d'audit spécifiques pour certains sites et nous traitons tous les webmasters de la même façon. Quand nous faisons des interventions lors de "site clinics" (*NDLR : comme lors du dernier SEO Campus en France*), nous le faisons dans le cadre de nos recommandations de qualité et pour montrer les "best practices" à suivre en termes de référencement naturel et d'optimisation. Quand nous publions des ressources en ligne comme des articles ou des vidéos dans notre Centre d'aide ou sur nos pages d'aide en ligne, elles sont disponibles à tous les possesseurs de site web. Il en est de même pour le positionnement des pages dans nos résultats, qui est indépendant, par exemple, de nos prestations de liens publicitaires. Un site qui est un excellent client en termes d'Adwords ne sera jamais mieux classé qu'un autre sur ce critère dans les résultats "organiques". Ce point, évident pour nous, doit souvent être rappelé...

Notre opinion : nous avons vu dans le passé plusieurs cas de sites web ayant comme volonté première de faire un travail "propre" en termes de référencement naturel... et finalement s'orienter vers des pratiques "limites", voire "black hat" devant l'impossibilité notoire d'obtenir des résultats qui soient au moins "acceptables" sur Google (et les webmasters voyant à quel point leurs concurrents leur passaient devant en spammant le moteur en toute impunité...). Nous restons

persuadé que, dans ce type de cas, une aide d'un expert SEO interne à Google aurait réglé bien des problèmes pour certains cas techniquement très complexes...

- Matt Cutts a expliqué, lors d'une interview réalisé par Eric Enge (<http://www.stonetemple.com/articles/interview-matt-cutts-012510.shtml>) que les redirections 301 pouvaient engendrer une perte de PageRank. Pouvez-vous nous en dire plus à ce sujet ?

Une très très faible partie du PageRank peut être perdue par ce biais. L'idée étant que si aucun PageRank n'était perdu, certaines personnes seraient tentées de faire certaines "astuces" plus ou moins "propres" avec les redirections 301 plutôt que d'utiliser des liens classiques. Les redirections 301 sont toujours de bonnes solutions et nous encourageons les webmasters à les utiliser quand une page a changé d'adresse de façon permanente.

Notez par ailleurs que, parmi tous les moyens d'indiquer qu'une page est un duplicate d'une autre, la redirection 301 est la voie qui est la plus transparente (celle qui fait perdre le moins de PageRank et d'autorité).

Notre opinion : pas de commentaire particulier, la redirection 301 est et restera un système indispensable en termes de référencement naturel. Et, bonne nouvelle, cela n'altère pas, ou si peu, la mise à jour d'URL...

- Parfois, on entend dire que le TBPR (ToolBar PageRank, c'est-à-dire l'indicateur proposé par la barre d'outils de Google) va disparaître, pour plusieurs raisons (marché parallèle des liens vendus au prorata du PR, par exemple). Pour nous, il semble évident que le PageRank lui-même ne peut disparaître, mais que cela pourrait être le cas de son "image" dans la toolbar. Pouvez-vous nous en dire plus ?

Bien que personnellement cela ne me chagrinerait pas que le TBPR disparaisse, il est important de se souvenir que la "webspam team" n'est qu'une partie d'une grande compagnie. Le fait est que de nombreux utilisateurs du Web et de notre barre d'outils aiment bien regarder l'indicateur de PageRank dans cette dernière pour avoir une idée de la confiance qu'ils peuvent avoir dans une page particulière. Personne ne sait ce que le futur nous réserve, mais il y a pour le moment autant d'arguments pour le supprimer que pour le garder...



Le "ToolBar PageRank", bientôt disparu ?

Notre opinion : en gros, si ça ne tenait qu'à la "webspam team", le TBPR n'existerait plus depuis longtemps. Mais d'autres services, chez Google, désirent le garder. Honnêtement, vous connaissez beaucoup d'utilisateurs qui regardent le PR dans la barre d'outils pour savoir s'ils peuvent avoir confiance dans un site ? A notre avis, le TBPR est plus un outil de communication pour Google (il montre qu'il a révolutionné le monde des moteurs de recherche à la fin de années 90 avec cette

innovation) et c'est pour cela qu'il est encore présent dans la toolbar... Jusqu'à quand ?? Ceci dit, le "vrai" PageRank a encore de beaux jours devant lui...

- Notre point de vue est que le concept de "positionnement" devient peu à peu obsolète, car les SERPs sont de plus en plus personnalisées à l'utilisateur. Ainsi, il peut sembler plus intéressant de se concentrer sur l'analyse de la qualité du trafic généré par les moteurs de recherche. Et l'utilisation d'outils comme Google Analytics peut souvent s'avérer plus efficace que des outils de suivi du positionnement pour mesurer la qualité d'une stratégie SEO. Quelle est votre opinion à ce sujet ?

Les moteurs sont en constante évolution et nous offrons aux utilisateurs du moteur Google des expériences diverses avec la recherche personnalisée mais aussi avec la recherche universelle. La recherche personnalisée permet de fournir des résultats dépendant de l'historique de recherche de l'utilisateur. En termes de référencement naturel, il est important de regarder les SERPs (pages de résultats du moteur) de façon globale, par exemple avec des médias et des supports qu'on peut utiliser pour gagner de la visibilité comme la recherche temps réel, les images, Google Adresses, etc.

Se positionner sur une phrase isolée est moins important que tous vos positionnements, et le positionnement est moins important que le taux de clic sur les liens ainsi positionnés. De même, le taux de clic n'est pas aussi important que le taux de conversion. Vous pouvez réellement vouloir maximiser vos objectifs (donc vos profits) sur la qualité du trafic issu du "search", pas nécessairement en tentant juste d'obtenir du trafic en quantité, mais non qualifié ou émanant du simple positionnement sur quelques requêtes isolées...

Notre opinion : nous sommes d'accord avec Google : le positionnement n'est important que s'il génère un trafic de qualité. Et pour cela, il nous semble bien plus intéressant de mesurer le taux de transformation (combien de visites les moteurs de recherche envoient-ils sur mon site et que font ces visiteurs sur mes pages ?) qu'un strict positionnement qui ne veut pas (plus) dire grand chose...

Merci Matt, John, Nathalie, Susan et Anouar pour ces réponses !

Interview réalisée par mail par Olivier Andrieu, éditeur du site Abondance.com, en juillet et août 2010.

Réagissez à cet article sur le blog des abonnés d'Abondance :
<http://blog-abonnes.abondance.com/2010/09/10-questions-la-search-quality-team-de.html>