

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Qu'on les appelle "robots", "spiders", "bots" ou "crawlers", les logiciels capables d'explorer les sites web pour mettre à jour leurs index ont toujours été utilisés par les moteurs de recherche. Mais les techniques ont changé depuis bientôt 20 ans. Entre les "traitements par lots" du début et le système mis en place par Google pour sa dernière mise à jour d'indexation Caffeine, il y a un monde que cet article se propose de vous faire découvrir. De quoi apporter quelques solutions aux webmasters qui ont connu quelques soucis d'indexation de leurs pages dans Gogle ces derniers mois...

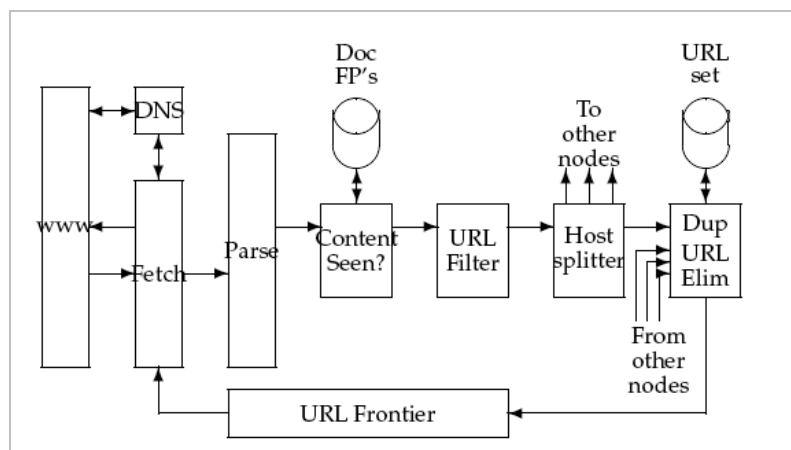
Les robots d'exploration des moteurs de recherche, que l'on appelle plus volontiers "crawlers", "spiders" ou "bots" sont l'un des composants-clé des moteurs de recherche. Malgré leur simplicité conceptuelle, la mise en œuvre des robots d'exploration de haute performance représente un défi d'ingénierie d'envergure en raison de la taille gigantesque du Web, et de l'évolution toujours plus rapide des pages.

Afin d'explorer une fraction substantielle du Web dans un laps de temps raisonnable, les robots d'exploration doivent télécharger des milliers de pages par seconde, et sont généralement répartis sur des dizaines ou des centaines d'ordinateurs. Leurs deux principales structures de données - la "frontière" (liste d'URL non encore explorées) et l'ensemble des URL déjà découverts - sont trop importantes pour entrer dans la mémoire principale, quels que soient les astuces trouvées pour "compresser" les URL et les données. La volonté de donner la priorité de l'exploration vers des pages de haute qualité et de préserver la fraîcheur de l'index impose des défis d'ingénierie supplémentaires.

Enfin, à ces défis techniques s'ajoute une contrainte : celle de rester "poli" envers les serveurs web, c'est à dire d'éviter de perturber et/ou de surcharger ceux-ci.

Dans le même temps, la course à la fraîcheur et à la rapidité est lancée dans le petit monde des moteurs de recherche : après la recherche "temps réel", on parle maintenant de recherche "instantanée".

Dans cet article, nous allons nous intéresser à la manière dont Google semble intégrer cette contrainte de "politesse" dans sa manière de crawler nos sites webs, et nous demander quelles évolutions Caffeine a réellement apporté dans les méthodes de crawl de Google ?



Architecture simplifiée d'un crawler distribué

La nécessité d'un crawl "respectueux" des serveurs Web

Il y a une quinzaine d'années, il n'était pas rare de voir les webmasters "bannir" les IP des crawlers des moteurs de recherche pour économiser leur bande passante. Google a connu cette mésaventure dans ses premières heures et à l'époque de leur prototype : Backrub. L'université de Stanford, par exemple, avait en effet exclu le robot inventé par ses deux étudiants, et le lancement du moteur avait déclenché un grand nombre d'appels téléphoniques de webmasters courroucés par le comportement de Googlebot.

Il existe deux types de mesure permettant d'éviter les désagréments provoqués par les robots d'exploration : les mesures explicites et implicites.

Les **mesures explicites** permettent aux webmasters de déclarer quels robots sont autorisés à crawler les pages de leur site, et quelles pages sont concernées. Les webmasters ont à leur disposition trois outils : le fichier robots.txt, la balise meta robots, et la directive x-robots-tag.

Les **mesures implicites** consistent à programmer le crawler pour éviter trois types d'inconvénients :

- la consommation d'une part trop importante de la bande passante ;
- la surcharge du serveur.

Dans les premiers temps, les mesures implicites ont consisté à veiller à éviter que le crawler ne demande des pages à un même serveur à un rythme trop important. Evidemment, cela limite considérablement la quantité de pages que l'on peut télécharger par seconde.

La solution consiste donc à chercher à évaluer si un serveur est capable de délivrer plus de pages qu'un autre, et de détecter la présence d'un goulet d'étranglement (temporaire ou permanent). Mais trouver une méthode pour détecter ces "*bottlenecks*" a longtemps été une difficulté sur laquelle ont buté de nombreux spécialistes des outils de recherche.

Google et le polite crawling

Dans un brevet publié récemment (voir dans la bibliographie), Google décrit son architecture de crawl (pré-caffeine) et explique comment ils implémentent les critères de politesse dans leur comportement de crawl.

Les auteurs du brevet décrivent notamment le phénomène de "*load capacity starvation*" en ces termes :

"La capacité d'un serveur web hôte à supporter la charge est souvent limitée par la configuration matérielle du serveur web. Lorsque les demandes simultanées des divers robots d'exploration sont au-dessus de la capacité de charge maximale qu'un serveur web peut supporter, il est presque certain que certains des robots concurrents subiront des lenteurs dans les réponses du serveur web, et que certaines requêtes pourront même échouer.

Un tel phénomène est parfois appelé "load capacity starvation" (famine de capacité de charge) pour un crawler web. La load capacity starvation empêche les robots d'exploration de récupérer des documents sur un site web et de les transmettre à un indexeur en temps opportun, ce qui nuit à la fois à l'éditeur du contenu et à la fraîcheur des résultats générés par le moteur de recherche."

Notons au passage, que jusqu'à Caffeine, Google avait multiplié les crawlers spécialisés (pour Adsense, adwords, news, images etc...), et que les crawlers concurrents dont on parle ici sont les différents crawlers de Google.

Comment les équipes de Google ont-elles résolu le problème ? Tout simplement en appliquant une règle simple : lors du crawl, un serveur (**processeur de charge**) attribue un "bail" pour une période de temps déterminée à la paire "crawler/hôte web" pour crawler une ou plusieurs pages. Si le crawler réussit à crawler les pages dans le temps imparti, alors le bail est prématurément interrompu et la capacité de charge libérée pour d'autres crawlers. Si le crawler ne parvient pas à crawler tous les documents, un nouveau bail est demandé au processeur de charge qui attribuera (ou pas) un nouvelle part de la capacité totale de charge. Si par contre aucune requête n'est faite au processeur de charge quand le bail expire, la capacité de charge est libérée pour les autres crawlers.

Par contre, ce que n'indique pas le brevet c'est comment est évaluée la capacité maximum de charge de l'hôte, c'est-à-dire le nombre d'URL par seconde que le crawler peut télécharger sans surcharger le serveur. Il est clair que cette grandeur est évaluée automatiquement, mais

comment ? On en est réduit à des hypothèses. Le moteur peut se baser sur deux indices pour détecter une surcharge du serveur :

- une hausse du temps moyen de téléchargement ;
- une augmentation des erreurs renvoyées par le serveur (erreurs 500, timeouts=).

La hausse du temps moyen de téléchargement permet de détecter soit une saturation de bande passante, soit du serveur web lui-même...

L'augmentation du nombre des erreurs révèle, au delà de la saturation de l'hôte, la saturation d'autres composants comme la base de données ou l'applicatif web.

L'évolution des crawlers de Google

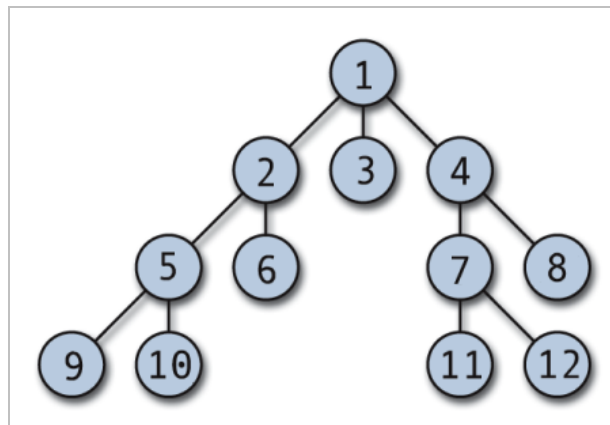
Le comportement des crawlers de Google a fortement évolué au fil du temps, et vient encore de changer radicalement cette année. Revenons tout d'abord sur l'historique de cette évolution.

La préhistoire des robots : le crawl par lots

Les premiers crawlers inventés fonctionnaient "par lots". Le principe était simple :

1. On part d'une liste d'URL à crawler (baptisée : url seeds. Disons 1000 à titre d'exemple) ;
2. On télécharge les URL en question (de la 1ère à la 1000e). Le crawl du lot est terminé ;
3. On découvre d'autres URL dans les pages crawlées (mettons 10000 nouvelles URL) ;
4. Les 10000 URL sont ajoutées au lot ;
5. Et on relance le crawl du lot composé cette fois de 10000 URL.

Cette méthode de crawl simplifiée utilisait un ordre de crawl baptisé BFS (*Breadth First Search*) par opposition à DFS (*Depth First Search*). La méthode BFS, combinée avec une règle de type "un seul hôte interrogé à la fois" a tendance à générer des visites plus espacées sur les serveurs web tout en garantissant un crawl efficace et rapide.



La méthode BFS en action : les nœuds sont explorés dans l'ordre des numéros

Le problème avec cette méthode, c'est qu'elle était très lente (de 5 à 10 jours pour Google à l'époque des "Google Dance", c'est-à-dire avant 2003), et que le crawler passait à côté de nombreuses pages éphémères, ou ne découvrait des pages que 3 semaines après leur mise en ligne !

Le crawl incrémental sur un Web "ouvert"

Google a probablement changé son infrastructure et ses méthodes de crawl à partir de la fin du printemps 2003 (update "Dominic"), pour progressivement passer à un système de crawl incrémental.

Avec le crawl incrémental, le crawl n'est jamais considéré comme achevé, mais se déroule de manière continue et se renouvelle perpétuellement. Les nouvelles URL découvertes sont ajoutées à la liste "frontière" (les nouvelles URL à découvrir) et crawlées dès que possible. Les URL déjà connues sont recrawlées à intervalle régulier.

Un **ordonnanceur** ("scheduler" en anglais) décide de crawler certaines nouvelles URL en priorité : il se base sur des critères comme l'importance de la page (pagerank) ou certains critères de qualité du contenu.

L'ordonnanceur décide également de la fréquence de recrawl des pages existantes en fonction de l'historique de changement de celles-ci, avec pour objectif de préserver un index aussi frais que possible.

Le crawl "Multicouches" : la version "pré Caffeine"

Avec le temps, les ingénieurs de Google ont été confrontés à des défis techniques de plus en plus pointus. L'augmentation de la taille du web, la volonté d'avoir un index plus important chez Google et l'ajout progressif de multiples crawlers spécialisés (pour les news, les images, les vidéos) pour aboutir aux fonctionnalités de la recherche universelle, ont réclamé une approche plus efficace lors des phases de crawl et d'indexation.

Google s'est tourné apparemment vers une méthode de crawl en plusieurs couches d'URL, décrite dans un brevet publié en 2007. La méthode décrite dans le brevet semble assez proche du comportement de Google avant Caffeine, elle mérite donc d'être détaillée, surtout qu'une grande partie de ces fonctionnalités sont probablement encore présente dans la version "Caffeine" du crawler.

La Couche de base

La couche de base est composée d'une séquence de segments. Chacun de ces segments peut comporter plus de deux cents millions d'URL. Une fois assemblés, les segments représentent un pourcentage important des URL présentes sur Internet. Les segments sont explorés périodiquement.

La Couche "Crawl Quotidien"

En plus des segments de la couche de base, il existe une couche d'exploration quotidienne. Les URL de cette couche d'exploration quotidienne sont explorés plus fréquemment que les URL de segments. Ceux-ci peuvent également inclure des URL "haute priorité" qui sont découverts par le système pendant un crawl en cours.

La couche Real-Time

La couche "temps réel" est composée d'URL qui doivent être explorés à plusieurs reprises au cours d'une époque donnée (par exemple, plusieurs fois par jour). Certaines URL dans cette couche sont explorées toutes les minutes ou presque. La couche temps réel comprend également les URL nouvellement découvertes qui n'ont pas été explorées, mais doivent être explorées dès que possible.

Un même robot pour toutes les couches

Les URL dans les différentes couches sont explorées par le même crawler, mais les résultats de l'exploration sont placés dans des index qui correspondent aux différentes couches définies plus haut. Un programme d'ordonnement du crawl définit le comportement du crawler en se basant sur l'historique de la fréquence de changement du contenu des pages web et sur une mesure de l'importance de l'URL (traduisez : sur une note de type Pagerank).

Passage d'une couche à une autre

Certaines URL peuvent être déplacés d'une couche à l'autre en se basant sur les logs qui indiquent à quelle fréquence le contenu associé à l'URL change, ainsi que différents rangs de page Web qui sont définies.

Le choix de la couche où il faut placer une URL est donnée par un calcul du score de type :

$\text{Score quotidien} = [\text{page rank}] \text{ au carré} * \text{fréquence de changement de l'URL}$
--

Mais qu'y a-t-il derrière Caffeine ?

Google a annoncé en septembre 2009 qu'ils allaient changer leur infrastructure (baptisée jusque là TeraGoogle, surnommée Bigdaddy par le forum WebmasterWorld) et se tourner vers une nouvelle version avec de meilleures capacités de crawl et d'indexation. Cette nouvelle infrastructure, baptisée Caffeine, a effectivement été progressivement testée et mise en place à partir de la fin de l'année 2009 et surtout au cours du premier semestre 2010. Google a annoncé officiellement la bascule vers Caffeine le 8 juin 2010.

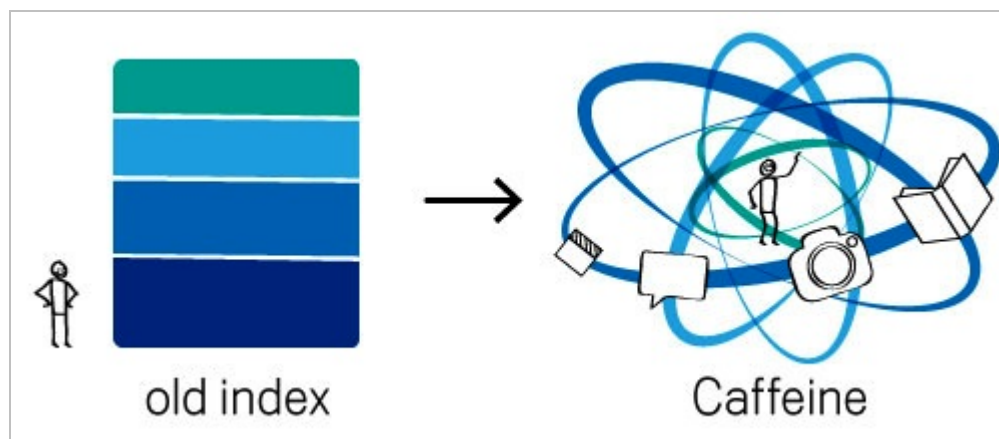
Cette nouvelle infrastructure est en fait composée de plusieurs éléments, comme une nouvelle version de l'"OS" de Google (GFS2, GFS = *Google File System*), de BigTable (= la couche de stockage des données, la nouvelle version permettant de manipuler des paquets de données avec une granularité plus faible), mais aussi une nouvelle infrastructure de crawl et d'indexation.

Caffeine : une nouvelle manière de crawler le web

On sait assez peu de choses sur les nouvelles méthodes de crawl et d'indexation utilisées par Google. Mais l'observation du comportement effectif des crawlers permet de trier parmi les différentes hypothèses et de choisir parmi les plus probables.

Google a déclaré avoir abandonné son système "multicouches" pour une exploration du web par petits bouts. Cela sous entend deux modifications majeures :

- Google n'a plus besoin de calculer des critères sur une couche entière pour déterminer si la page est importante ou non, doit être recrawlée ou non ;
- la stratégie de crawl de Google est en rupture par rapport à la stratégie *Breadth First Search* et ses avatars.



En fait il semble que Google ait remplacé (partiellement) les critères traditionnels par une approche statistique. Le billet annonçant le lancement de Caffeine a été signé par Carrie Grimes. Carrie Grimes est une brillante statisticienne qui a collaboré avec Google en tant que chercheur à l'Université de Stanford, puis a été embauchée par la firme de Mountain View. Elle est l'auteur d'une méthode capable de prédire, à partir d'une analyse statistique du passé d'un groupe de pages observé par le crawler, la probabilité d'un changement. Ces "prédictions" permettent donc de déterminer le comportement de recrawl du moteur de recherche.

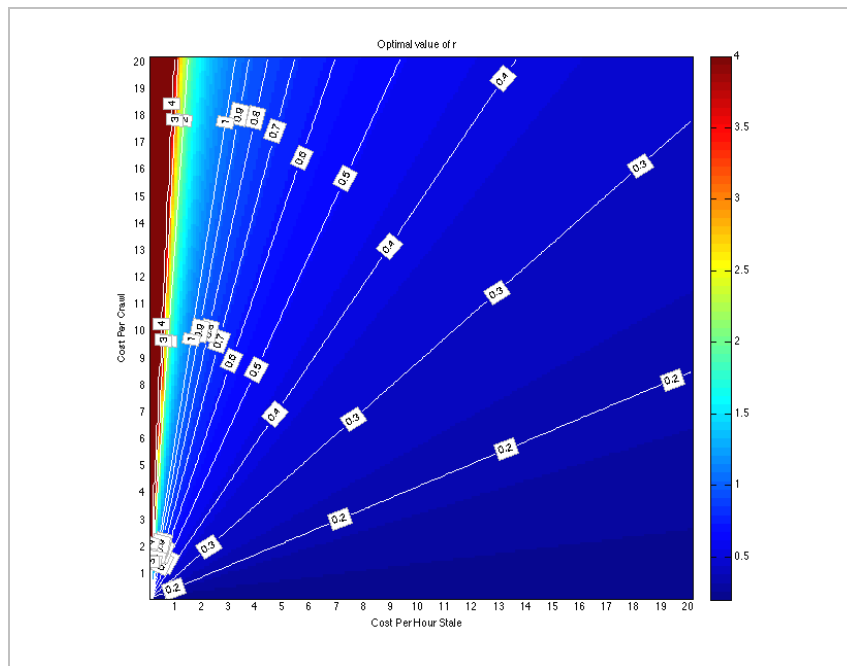
De nouvelles fonctions d'estimation de la nécessité de recrawl

La plupart des méthodes déterminant le "recrawl" d'une page se basaient jusqu'à maintenant sur les travaux d'Hector Garcia Molina qui avait introduit deux familles de critères pour décider si une page devait être crawlée à nouveau ou pas :

- des règles basées sur l'importance de la page (et en particulier son Pagerank) ;
- des règles basées sur la fréquence de changement des pages.

La plupart des personnes qui ont cherché à théoriser la fréquence de changement des pages sont parties du principe qu'un modèle de type "loi de Poisson" était applicable. Hector Garcia Molina avait développé une fonction d'estimation de cette fréquence de changement en fonction de l'observation des changements sur les pages baptisé "MLE" (*Maximum Likelihood Estimator*).

Carrie Grimes a enrichi cette approche avec de nouvelles fonctions d'évaluation, notamment du "coût" d'un recrawl et du "risque" qu'une page devienne obsolète si elle n'est pas recrawlée à temps. Et elle a surtout introduit une méthode qui permet de déterminer la fréquence de recrawl des nouvelles pages découvertes d'une manière efficace, la fréquence de recrawl évoluant rapidement vers une estimation correcte.



Graphique analysant la relation entre coût du crawl et coût de l'obsolescence de la base pour différentes valeurs de la fréquence de recrawl

Une nouvelle façon d'indexer

Avec Caffeine, il n'est plus utile d'attendre que l'ensemble de la couche d'URL ait été crawlée pour lancer le calcul des signaux et des critères de classement. Les pages sont traitées par petits groupes d'URL, si on en croit le communiqué de Google, et directement traitées pour être indexées.

Cela signifie que Google ne calcule plus certains critères sur des matrices d'URL complètes, mais sur des collections locales de pages : il est donc probable que certains signaux de type Pagerank sont estimés, pour être ensuite affinés de temps à autre.

L'avantage de cette méthode c'est que l'on peut réduire considérablement l'intervalle de temps entre le crawl initial d'une page, et sa disponibilité en tant que résultat dans une page du moteur de recherche.

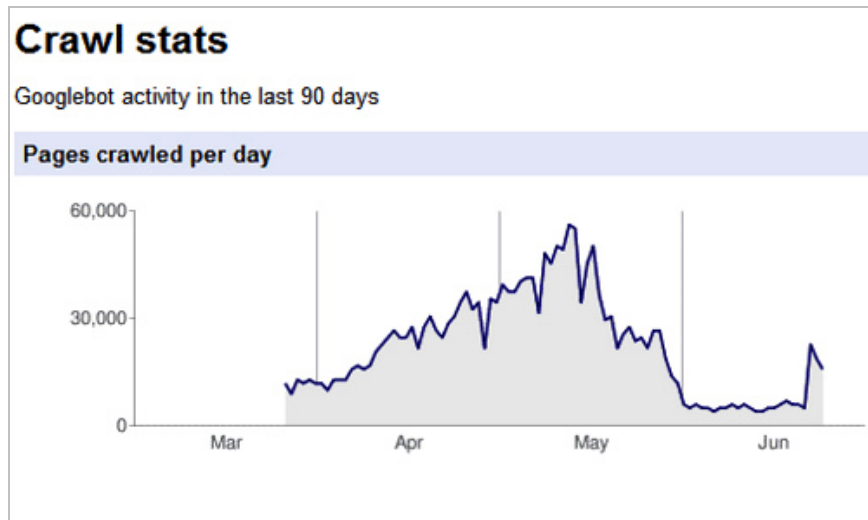
Quelles conséquences pour le référencement ?

Les conséquences de ces changements sont assez importants pour être sérieusement pris en compte par les webmasters. Dès qu'un site atteint une certaine taille, une partie des pages ne sont pas crawlées, et *a fortiori* indexées par le moteur. Connaître les critères qui régissent le crawl et l'indexation d'une page et savoir agir sur ceux ci font partie des éléments clés d'un bon référencement (une page non indexée n'a aucune chance d'apporter du trafic issu d'un moteur de recherche).

Caffeine n'a probablement pas révolutionné ces critères, mais il est clair que la liste des pages crawlées, et la liste des pages indexées a changé.

Par ailleurs, la fréquence de recrawl des pages et la fraîcheur de celles ci est devenu un élément important à surveiller. L'utilisation de nouvelles fonctions d'évaluation a des effets de

bord incontestables : si certaines pages ne sont pas recrawlées souvent, alors qu'elles contiennent des liens vers des pages "fraîches", vous avez toutes les chances de voir une partie de votre contenu ne jamais être découvert par le moteur.



Activité typique du passage à Caffeine sur un site.

Certains sites ont connu au contraire un pic de crawl dans les derniers jours d'avril, mais le plus grand nombre a connu une chute de crawl importante en juin.

Comment détecter ce problème ?

Plusieurs méthodes simples peuvent être utilisées pour détecter les problèmes de crawl et d'indexation :

- la liste des pages recevant des backlinks communiquée dans le compte GWT (Google Webmaster Tools) est une indication des URL connues par Google. Attention : cette liste indique des URL découvertes par Google, mais qui peuvent ne pas être crawlées, ni indexées. La présence ou l'absence de syntaxes d'URL ou de groupes de pages permettent de détecter d'éventuels problèmes.

- l'analyse des logs serveur permet de suivre la liste des pages crawlées et la fréquence de recrawl. Pour des petits sites, l'utilisation de scripts spécialisés permet un accès direct à ces informations.

- les courbes communiquées dans GWT (nombre de pages crawlées /jour, nombre de ko téléchargés par jour, temps moyen de téléchargement) permettent de détecter un problème technique qui peut gêner le crawl.

- la commande "site:sousdomaine.nomdedomaine.tld/rep" donne une estimation du nombre de pages dans l'index. Il faut rappeler ici (beaucoup de webmasters l'ignorent) que sa syntaxe a évolué, et que l'on peut avoir des résultats par sous domaines et par répertoire ou par racine d'url lorsqu'il y'a des paramètres: site:www.domaine.com/page?numero renverra toutes les pages commençant par page, comme www.domaine.com/page?numero=124.

On peut utilement croiser la commande "site:" avec les options permettant de sélectionner les pages datant de :

- moins de 24h
- moins d'une semaine
- moins d'un mois
- mois d'un an

La commande site: sans ces options donne des résultats assez changeants et en général difficile à interpréter. La commande site: croisée avec les options de récence donne des résultats plus précis. Une baisse du nombre des pages récentes, notamment du nombre de pages datant de moins d'une semaine doit vous alarmer.

Comment réagir si Caffeine ne crawle plus ou n'indexe plus vos pages ?

- un sitemap xml contenant les pages "oubliées" peut aider ;
- penchez-vous sur la structure de votre site : si les liens vers vos pages fraîches se situent sur des pages crawlées tous les mois, placez sur des pages plus hautes dans l'arborescence et/ou crawlées plus souvent
- demandez-vous pourquoi l'estimateur se trompe : si vos pages sont crawlées moins souvent que leur rythme d'évolution réel, c'est soit qu'elles sont jugées peu importantes, soit de mauvaise qualité, soit que leur évolution n'est pas détectée !

Et demain ?

On peut s'attendre dans les mois qui viennent à de nouvelles évolutions dans le comportement de crawl. L'avantage des méthodes statistiques, c'est qu'elles sont évolutives par essence. Il est possible à chaque instant de "recalibrer" le système, d'affiner les estimateurs, et surtout, d'analyser les rétroactions et de comprendre si le nouveau comportement de crawl est efficace ou non en termes d'efficacité, de fraîcheur et de qualité du contenu téléchargé. Après plusieurs mois d'accumulation de données et d'expérimentation, il est possible que le comportement de crawl observé aujourd'hui évolue vers quelque chose de différent. Pour le moment certains indices laissent penser que les choses ne sont pas encore "stabilisées".

Mais certains pensent déjà à l'étape suivante. L'une des solutions pour accélérer les choses serait d'arrêter de penser en mode "pull" (demander des données au serveur), pour évoluer vers des serveurs web capables de "push" (envoyer les données au moteur). C'est l'approche du protocole PubSubHubbub, mais ce dernier ne propose pas de solution permettant d'éviter le spamdexing. Certains spécialistes des outils de recherche expérimentent donc des solutions s'appuyant sur des agents déportés, des logiciels fournis par les moteurs, s'implantant localement au coeur des serveurs webs et facilitant la découverte des nouvelles pages.

On peut donc parier sur le fait que Caffeine n'est qu'une étape dans l'évolution des méthodes de crawl et d'indexation et que beaucoup de choses restent à imaginer et à inventer.

BIBLIOGRAPHIE

Si ce sujet vous intéresse, vous pouvez approfondir vos connaissances sur les crawlers et la problématique de la priorité de crawl et de recrawl en lisant les articles suivants :

Keeping a Search Engine Index Fresh: Risk and optimality in estimating refresh rates for web pages

<http://www.google.com/research/pubs/archive/34570.pdf>

D. Ford, C. Grimes, and E. Tassone, Google

Microscale evolution of web pages.

<http://www.google.com/research/pubs/archive/34428.pdf>

S. O'Brien and C. Grimes

In WWW '08: Proceedings of the 17th International World Wide Web Conference, 2008.

Detecting Near-Duplicates for Web Crawling

<http://research.compaq.com/SRC/mercator/papers/www10.ps>

Gurmeet Singh Manku, Google Inc, Arvind Jain, Google Inc, Anish Das Sarma, Stanford University

Breadth-First Search Crawling Yields High-Quality Pages

<http://research.compaq.com/SRC/mercator/papers/www10.ps>

Marc Najork, Janet L. Wiener, Compaq

Web Crawling

<http://research.microsoft.com/pubs/121136/1500000017.pdf>

Christopher Olston and Marc Najork

(une "somme" de 74 pages sur l'état de l'art sur les crawlers : incontournable)

Effective Page Refresh Policies For Web Crawlers

<http://oak.cs.ucla.edu/~cho/papers/cho-tods03.pdf>

Junghoo Cho, University of California, Los Angeles and Hector Garcia-Molina, Stanford University

Recrawl Scheduling Based on Information Longevity

<http://www-db.stanford.edu/~olston/publications/www08.pdf>

Christopher Olston, Yahoo! Research, Sandeep Pandey, Carnegie Mellon University

Estimating Frequency of Change

<http://www-db.stanford.edu/pub/papers/cho-freq.ps>

Junghoo Cho, University of California, and Hector Garcia-Molina, Stanford University

Efficient crawling through URL ordering.

<http://www-db.stanford.edu/~cho/papers/cho-order.pdf>

J. Cho, H. Garcia-Molina, and L. Page.

Computer Networks and ISDN Systems, 30(1-7):161-172, 1998.

Crawling a Country: Better Strategies than Breadth-First for Web Page Ordering

http://www.tejedoresdelweb.com/slides/Crawling_a_Country.pdf

Ricardo Baeza-Yates, Carlos Castillo, Mauricio Mar and Andrea Rodriguez

BREVETS

Anchor tag indexing in a web crawler system

<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetacgi/nph-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7,308,643.PN.&OS=pn/7,308,643&RS=PN/7,308,643>

[.643](http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetacgi/nph-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7,308,643.PN.&OS=pn/7,308,643&RS=PN/7,308,643)

Invented by Huican Zhu, Jeffrey Dean, Sanjay Ghemawat, Bwolen Po-Jen Yang, and Anurag Acharya

Assigned to Google, US Patent 7,308,643, Granted December 11, 2007, Filed July 3, 2003

Limiting requests by web crawlers to a web host

<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetacgi/nph-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7,774,782.PN.&OS=pn/7,774,782&RS=PN/7,774,782>

[.782](http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetacgi/nph-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7,774,782.PN.&OS=pn/7,774,782&RS=PN/7,774,782)

Invented by Catalin T. Popescu and Anurag Acharya

Assigned to Google, US Patent 7,774,782, Granted August 10, 2010, Filed: December 18, 2003

Philippe Yonnet, Global SEO Strategist, WEB DMUK (Londres) – Easyroommate / Vivastreet

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

<http://blog-abonnes.abondance.com/2010/09/google-cafeine-et-levolution-des.html>