

L'algorithme de Google est-il modifié sur les requêtes "larges" ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Les moteurs de recherche actuels doivent faire face à plusieurs défis majeurs et parmi ceux-ci se trouve la taille grandissante du web. Comment, dans ce cas, répondre avec la plus grande efficacité à des requêtes, larges, génériques ou ambiguës ? Faut-il utiliser le même algorithme pour une demande très précise et une autre beaucoup plus large ? Plusieurs méthodes sont possibles et sont décrites dans cet article, mais Google pourrait bien en utiliser une spécifique, expliquée dans un brevet datant de 2004 et étudié dans cet article...

L'un des défis les plus complexes que les moteurs comme Google ont eu à relever ces dernières années est certainement la croissance forte de la volumétrie des données qu'ils doivent manipuler. Deux phénomènes conjoints sont à l'oeuvre derrière cette explosion. Tout d'abord, le nombre de pages web ne cesse d'augmenter. Ensuite, avec son infrastructure Caffeine, Google s'est doté en 2010 des moyens pour crawler et indexer beaucoup plus de pages, et pour les crawler et les indexer plus rapidement !

L'ensemble a produit une nouvelle donne, qui a conduit Google à revoir son algorithme sur plusieurs aspects. L'index du moteur étant devenu plus "profond" (c'est à dire qu'une proportion plus importante du WWW a été crawlée et indexée), l'un des effets de bord attendu était la présence beaucoup plus visible et massive des sites "poids lourds", comportant des millions de pages, dans les pages de résultat sur les requêtes "longue traîne". La mise à jour dite "MayDay" le 1er mai 2010, contemporaine du déploiement complet de Caffeine, avait clairement pour objectif de traiter ce problème.

Un deuxième effet de bord a été de favoriser de nouvelles formes de spamdexing, comme les fermes de contenu, ou plus généralement, tous les sites créant des millions de pages à partir de bases de données, le plus souvent "scrapées" c'est à dire constituée en copiant et en analysant les données déjà disponibles sur le web. On peut considérer que Panda est une réponse à ce nouveau problème : les index des moteurs sont inondés de pages de contenu créées massivement en utilisant des méthodes et des techniques qui ne violent pas clairement les conditions d'utilisation des moteurs de recherche (mais qui posent des problèmes de qualité).

La qualité des résultats sur les requêtes génériques

Mais cette explosion de la volumétrie des pages à classer produit également d'autres effets de bord. En particulier, la qualité des résultats sur les requêtes "génériques" s'est sensiblement dégradée. Ce n'est pas un phénomène récent cette fois-ci, le problème du très grand volume de pages retournées par une requête générique a conduit l'algorithme à atteindre ses limites probablement dès 2005 pour les requêtes anglophones.

A partir de fin 2006, le déploiement de BigDaddy, l'infrastructure qui a précédé Caffeine, a produit le même phénomène d'accroissement de la taille de l'index, et le problème de la qualité des résultats sur les requêtes génériques s'est posée d'un coup avec une acuité nouvelle.

Dans cet article, nous allons étudier en détail l'une des réponses que Google a cherché à apporter à ce nouveau problème, et qu'ils ont détaillée dans un brevet dont le dépôt date... de 2004, et qui rend compte assez bien des évolutions remarquées ces dernières années sur ces requêtes.

Requête "générique", requête ambiguë ou requête "large" ?

Pour bien comprendre le problème, il faut déjà définir de manière plus précise ce que l'on entend par requête "générique". En réalité, il s'agit de requêtes composées d'un nombre réduit de termes courants, termes qui sont susceptibles d'être présents dans le contenu d'un nombre considérable de pages (des millions, quand ce n'est pas des milliards pour les cas les plus extrêmes). Le problème, c'est que le nombre de mots clés ne prédit pas toujours le nombre de pages que le moteur va retourner en réponse à la requête. Certaines requêtes à plusieurs mots renvoient beaucoup plus de résultats que des requêtes à un seul mot. La requête "psalliotte" ne renvoie que 15 000 résultats, alors que la requête "ipod nano 6g multi touch" renvoie 716 000 pages !

On préfère donc analyser en général le cas des requêtes larges ("*broad queries*"), plutôt que les requêtes génériques (appelées souvent "general queries" en anglais). Une requête large est, par définition, une requête qui renvoie un nombre très élevé de résultats (au delà d'un certain seuil donc), sans tenir compte de la longueur de la requête, ou de son caractère générique *a priori*.

De fait, ces requêtes larges, qu'elles relèvent de la "*short tail*" (courte queue, par opposition aux requêtes "*longues queues*") ou non, sont le plus souvent ambiguës. Mais cette ambiguïté porte soit sur le sens de la requête entrée par l'utilisateur, soit encore sur le type de résultats qu'il cherche (quel genre de sites répond le mieux à sa demande ?), soit enfin sur l'intention de l'utilisateur quand il tape une requête.

Le premier groupe de cas correspond à des requêtes de type "java" ou "jaguar" : le terme est ambigu, impossible de savoir quel sens donner à la requête, et il faut donc présenter des résultats qui relèvent des différentes familles de sens.



Sur la requête jaguar, Google ne sait pas quel est le sens de la requête : la page de résultats présente des pages sur les voitures Jaguar, et sur le félin d'Amérique du Sud.

Il faut remarquer que cette présence de pages sur le félin en tête des résultats ne découle pas du fonctionnement normal de l'algorithme : les pages sur les voitures jaguar sont beaucoup plus populaires, et devraient squatter les premières positions, comme sur la plupart des moteurs concurrents.

La seconde famille de cas correspond à des requêtes type "Victor Hugo". Que cherche-t-on exactement ? Une biographie de Victor Hugo ? Un site qui vend des livres de Victor Hugo ? Un poème de Victor Hugo ? Une bibliographie de Victor Hugo ? Une photo de Victor Hugo ? Impossible de deviner sans plus de contexte...

La troisième famille de cas correspond à des requêtes du type : "langage java". La requête est toujours aussi large : sur l'index français, elle renvoie plus d'un million de résultats. Le sens ne fait plus de doute : il s'agit là du langage informatique, pas de l'île d'Indonésie, ni de la danse. Mais quels résultats faut-il présenter en réponse à la requête ? Le site officiel du langage ? Des ressources sur le langage ? La page de Wikipedia sur le langage ?

Remarque : Google travaille de mieux en mieux sur la reconnaissance d'expressions ("*phrases*" en anglais, attention c'est un « faux ami »). Dans ce cas précis la requête "[langage java](#)" est reconnue comme une expression, et renvoie un million de résultats. La requête "[java langage](#)" en renvoie ... 5 fois plus.

Pourquoi ?

"[langage java](#)" est une expression, qui appelle en réponse des pages dans lesquelles les mots "[langage](#)" et "[java](#)" apparaissent dans cet ordre précis, et sans rien pour les séparer. Par contre "[java langage](#)" n'est pas une expression, mais une juxtaposition de mots clés. Les pages à présenter en réponse peuvent contenir "[java](#)" et "[langage](#)" dans n'importe quel ordre, les termes étant éventuellement séparés par d'autres mots dans les pages web. C'est l'une des améliorations apportée par l'indexation des phrases dans l'algorithme de Google.

Mais il existe aussi des cas de requêtes qui ne sont pas ambiguës, et qui renvoient des centaines de milliers de résultats. Ces requêtes restent larges. Par exemple la requête : "[acheter une voiture d'occasion pas cher](#)" n'est pas ambiguë sémantiquement, elle révèle une intention claire de l'utilisateur, et l'on devine le type de pages/sites qu'il faut présenter à l'utilisateur pour qu'il juge les résultats pertinents. Mais "[acheter une voiture d'occasion pas cher](#)" qui renvoie 4,5 millions de résultats est clairement à classer dans les requêtes "larges".

Les réponses apportées par les moteurs au problème des requêtes larges et ambiguës

Le problème des requêtes larges et ou ambiguës n'est pas nouveau, aussi les moteurs ont-ils inventé depuis longtemps des solutions pour faire face au problème posé par l'ambiguïté des requêtes.

Les solutions les plus courantes sont :

- le raffinement de requêtes ;
- la reformulation des requêtes ;
- la diversification des types de résultats présentés ;
- la détection de l'intention la plus probable ;
- la personnalisation des résultats.

Le raffinement de requête

Le raffinement des requêtes consiste à suggérer d'autres requêtes, plus précises et moins ambiguës à l'utilisateur. On passe donc d'une requête ambiguë à une requête claire, explicite. Les principaux moteurs (Yahoo !, Google, MSN à l'époque) ont longtemps rechigné à adopter cette solution classique que l'on a donc trouvé plus facilement sur les moteurs alternatifs. Il faut dire que cette solution a un inconvénient majeur : elle oblige l'utilisateur à faire une action supplémentaire.

Mais depuis quelques années, on trouve des liens relevant clairement du raffinement des requêtes sur Yahoo !, et de façon un peu moins « pure » sur Google et Bing.



The screenshot shows the Yahoo! search interface. At the top, there's a navigation bar with links like 'Web', 'Images', 'Vidéo', 'Local', 'Shopping', 'Actualités', and 'Plus'. The search bar contains the text 'jaguar'. Below the search bar, there's a dropdown menu with suggestions: 'jaguar xj', 'jaguar france', 'jaguar occasion', 'jaguar xk8', 'jaguar xf', 'Affinez vos résultats : Jaguar Type', 'Jaguar XJ', 'Jaguar XF', 'véhicule', 'Jaguar occasion', 'Jaguar XKR', 'félins', and 'Panthera onca'. Below the search bar, there are search filters: 'Recherche : sur tout le Web', 'en français', and 'en France'. On the left, there's a 'Filtre adulte - Désactivé' button and a count of '413 000 000 résultats pour jaguar:'. The main results area shows 'cherche Jaguar' with a sponsored link 'Trouvez votre voiture d'occasion Parmi toutes nos annonces . Ww.annoncesjaunes.fr/auto' and another sponsored link 'Jaguar xj sur ArgusAuto Découvrez les annonces occasions Jaguar xj sur Argusauto.com!'. On the right, there's another sponsored link 'Jaguar Trouvez votre Jaguar d'occasion Au meilleur prix sur 321auto.com www.321auto.com/jaguar-occasion'.

Raffinement de requêtes sur Yahoo ! Le moteur suggère des requêtes plus précises autour des automobiles « jaguar », mais aussi la requête « Panthera onca » qui renverra à coup sûr uniquement des pages de zoologie sur le félin d'Amérique du Sud.

La reformulation des requêtes

Avec le temps, les moteurs se sont rendu compte que le problème de l'ambiguïté des requêtes génériques serait résolu... si l'on apprenait aux utilisateurs à taper des requêtes « claires ». Peu à peu, les internautes ont appris d'ailleurs à taper des requêtes plus longues pour obtenir des résultats plus précis. Mais force est de constater que la plupart d'entre eux ne parviennent pas à entrer des requêtes bien formulées du premier coup.

Le premier problème à résoudre est celui des requêtes mal orthographiées : dans ce cas la reformulation des requêtes consiste à présenter les résultats avec la bonne graphie en regard des résultats avec la mauvaise orthographe et/ou de proposer un lien vers la requête correcte probable (le fameux « Did you mean... » sur le moteur Google).

Mais la solution la plus élégante à ce problème est celle apportée par les suggestions de requêtes (le système « d'autocomplétion »). Ce système a clairement pour objectif d'aider l'utilisateur à formuler sans erreur de graphie des requêtes plus claires. Les suggestions de requêtes ont été complétées sur Google par « Google Instant » qui permet en plus d'avoir un aperçu en temps réel des résultats obtenus au fur et à mesure qu'on tape.

La diversification des types de résultats présentés

Malgré tous les efforts accomplis pour aider l'utilisateur à formuler une requête « claire », la requête peut rester ambiguë. Que cherche-t-on exactement si l'on tape la requête « Erin Brockovich » ? Des informations sur la vraie « Erin Brockovich », sur le film éponyme qui retrace sa vie, ou sur le personnage dans le film interprété par Julia Roberts ? Quel type de documents veut-on trouver d'ailleurs ? Des images, des vidéos, des pages de news, des critiques de films, un article de Wikipedia ?

The screenshot shows a Google search for "erin brockovich". The search bar contains the text "erin brockovich" and a search button. Below the search bar, it says "Environ 2 830 000 résultats (0,11 secondes)". To the right, there are links for "Google.com in English" and "Recherche avancée".

The search results are categorized into four groups, each indicated by a red arrow pointing to the right:

- Images mélangées:** A row of seven small image thumbnails showing Erin Brockovich in various settings.
- Personne:** A link to the Wikipedia page for "Erin Brockovich (personne)". The snippet below the link reads: "Erin Brockovich-Ellis (née Erin L. E. Pattee, le 22 juin 1960, à Lawrence, Kansas) est une autodidacte, devenue adjointe juridique et militante de ...".
- Personnage:** A link to the Wikipedia page for "Erin Brockovich, seule contre tous". The snippet below the link reads: "Erin Brockovich, seule contre tous (Erin Brockovich) est un film américain ...".
- Personnage:** A link to the AlloCiné page for the movie "Erin Brockovich, seule contre tous (1999)". The snippet below the link reads: "Mère élevant seule ses trois enfants, Erin Brockovich n'avait vraiment pas besoin d'un accident de voiture. D'autant que le responsable sort du tribunal financièrement indemne. Obligée... Réalisée par Steven Soderbergh. Avec Julia Roberts, Albert Finney, Aaron Eckhart. Casting - D'infos - Spectateurs - 12".

At the bottom, there is a link to the "Official Website of Erin Brockovich" with the snippet: "Erin Brockovich - CONSUMER ADVOCATE: 19 years and still fighting!".

La seule solution dans ce cas est de s'assurer que la première page de résultats contient un « pot-pourri » de tous ces éléments : des pages parlant de la personne, du film, et du personnage. Et des news, des images, des vidéos, et des pages html.

Ce dernier « mélange » est assuré par la logique de « recherche universelle » que l'on retrouve à présent dans tous les moteurs majeurs : présenter dans les pages de résultats du moteur « textuel » des résultats issus de différents moteurs verticaux.

Mais créer de la diversité sémantique (comme dans l'exemple Jaguar cité plus haut) ou thématique (comme dans l'exemple Erin Brockovich ci-dessus) est loin d'être facile. Il faut en effet :

- étiqueter (catégoriser) les pages en fonction de leur nature, de leur sens, de leur thématique ;
- déterminer quelles pages présenter pour une famille de requêtes données ;
- et servir une liste de résultats « diversifiées » à la vitesse de l'éclair.

C'est en fait un véritable challenge, et les moteurs comme Google ne gèrent correctement cette logique de diversité que depuis quelques années seulement.

Remarque : Amit Singhal, le « patron » de l'algorithme de Google, dans une interview accordée au New York Times, avait révélé quel nom était donné en interne aux requêtes identifiées comme appelant de la diversité : les requêtes QDD (*Query Deserves Diversity*) par symétrie avec les requêtes QDF (*Query Deserves Freshness*).

La détection de l'intention la plus probable

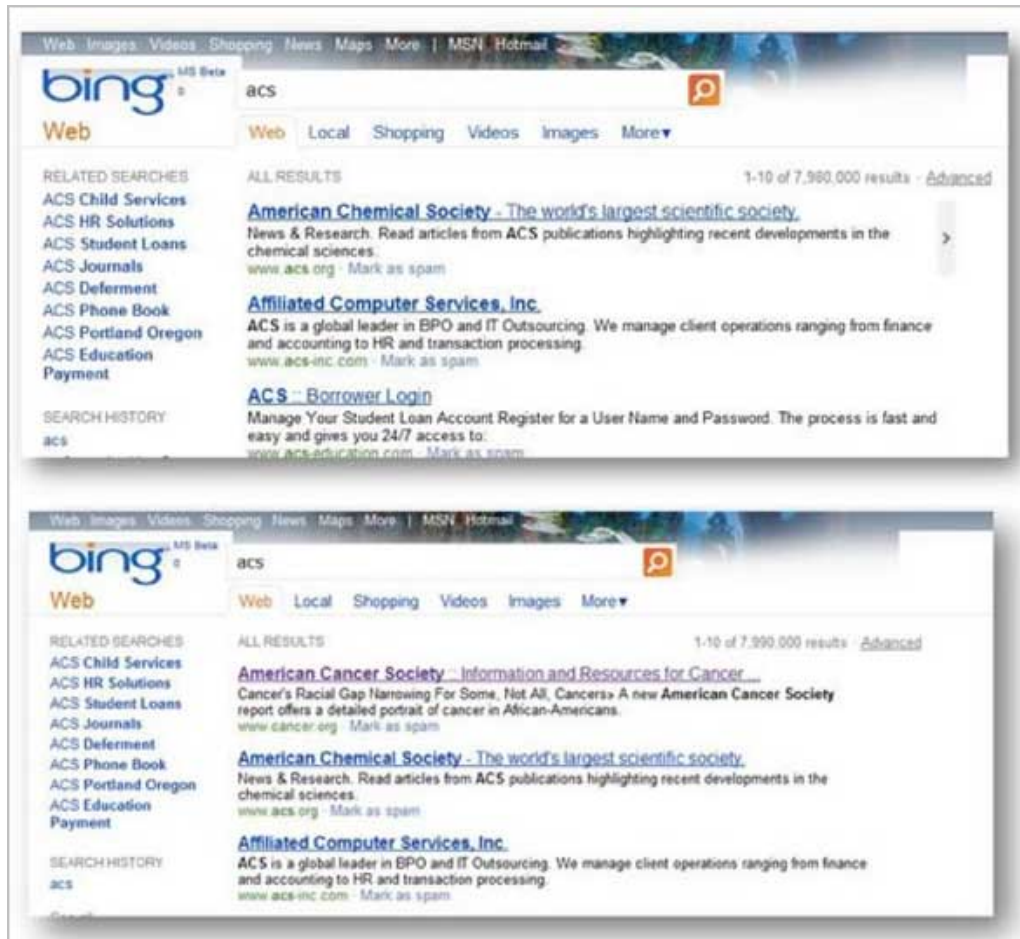
Dans certains cas, il est possible de « sélectionner » les résultats à présenter sur une requête ambiguë en devinant l'intention probable de l'utilisateur qui tape cette requête. L'analyse des logs de requête permet en effet de savoir ce que font les utilisateurs après avoir tapé une

requête ambiguë (la plupart du temps ils retapent d'autres mots clés et cherchent à raffiner leur requête).

Ces données sur les intentions probables des utilisateurs permettent, sur certaines familles de requêtes, de proposer directement des résultats filtrés selon l'objectif implicite détecté.

La personnalisation des résultats

Une autre solution est aussi de chercher à utiliser l'historique des recherches de l'utilisateur, ou sa géolocalisation, ou toute autre information de contexte, pour faire remonter les résultats les plus pertinents pour un utilisateur donné.



La personnalisation des résultats chez Bing : deux utilisateurs différents voient deux listes de résultats classés différemment pour la requête « ACS ».

Ces solutions ne sont pas universelles

Le problème avec toutes les solutions « classiques » présentées ici, c'est qu'elles ne sont pas universelles. Leur apport en termes de pertinence est incontestable, mais dans la plupart des cas, elles demandent que l'on ajoute énormément d'informations dans le système pour qu'elles fonctionnent.

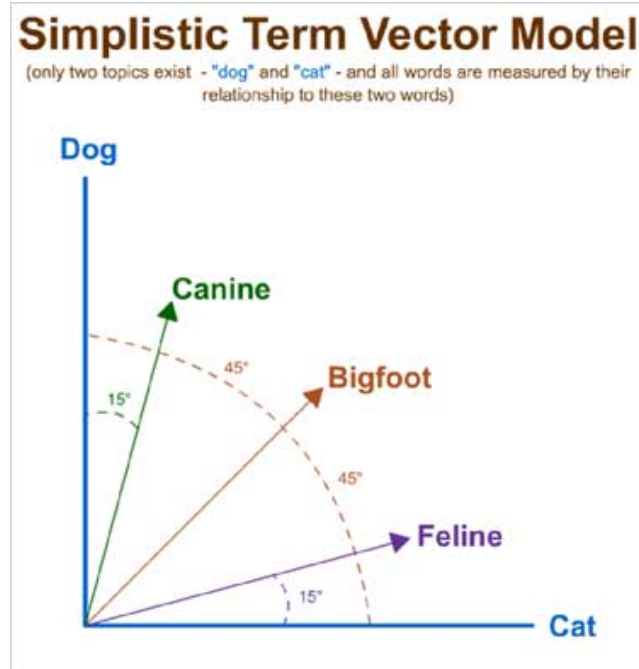
Même si ces informations sont calculées par un algorithme, jamais toutes les requêtes ne pourront être couvertes par ces solutions. Les moteurs ont donc aussi besoin de solutions plus universelles pour classer correctement les résultats sur des requêtes « larges ».

Comment Google classe-t-il ses pages habituellement ?

Pour comprendre le problème, il faut revenir aux mécanismes fondamentaux à l'œuvre dans les moteurs de recherche.

Les moteurs de type « Google » s'appuient sur un cocktail savant de signaux, mesurant :

- la proximité (similarité) sémantique de la page avec la requête.
- Et la popularité de la page (mais aussi des signaux de qualité dont la confiance, l'autorité, la topicalité...).



Le principe du classement par mesure de la distance angulaire entre des termes, projetés dans l'espace vectoriel des termes. Cette mesure de la similarité, baptisée Cosinus de Salton, n'est pas forcément utilisée dans l'algorithme de Google, mais l'algorithme contient forcément une mesure de similarité, indispensable pour renvoyer des résultats proches sémantiquement de la requête. Ill. Crédit seomoz.org

A ces signaux principaux se rajoutent probablement des signaux mesurant le comportement des utilisateurs sur les pages de résultat : la mesure du nombre de clics sur les URL présentées dans les résultats, et du taux de clics sur ces résultats. Ces données sont très utiles pour mesurer la qualité des réponses données par l'algorithme (retour de pertinence). Elles sont indispensables pour la personnalisation. Et jouent un rôle de plus en plus visible dans les algorithmes eux-mêmes.

Gérer des milliers d'ex aequo ou presque

La mesure de « proximité/similarité » entre la requête et la page joue un rôle majeur dans la pertinence. Les signaux associés à cette mesure ont donc un poids énorme dans l'algorithme. Mais sur une requête qui retourne 1 million de résultats, il est fréquent que le score de similarité soit quasiment identique, à quelques décimales près, avec des milliers d'autres résultats.

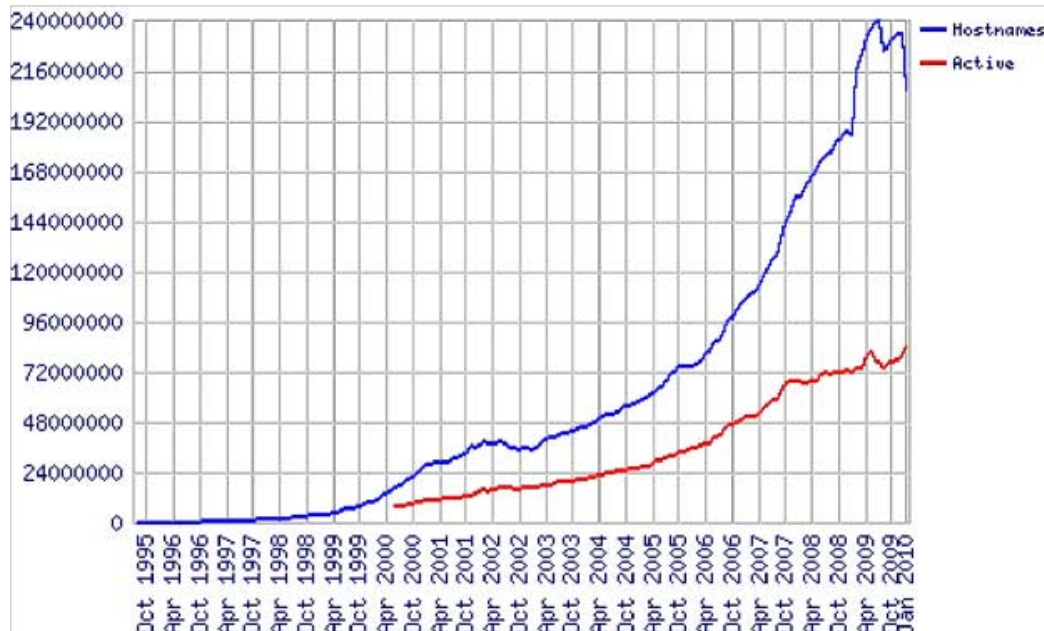
Le premier résultat retourné sur ce type de critères a donc autant de chances que le 1 000ème d'être considéré par l'utilisateur comme le résultat le plus pertinent !

C'est là qu'interviennent les mesures de l'importance de la page (c'est-à-dire le Pagerank en l'occurrence) : en redistribuant les classements entre ex-aequo, elles garantissent que les pages les plus importantes parmi les meilleurs résultats sortent en premier.

Par rapport aux critères d'importance, les critères de popularité de type « impressions » de résultats ou CTR servent également à redistribuer les résultats en faisant remonter les plus populaires (ce mécanisme est évident sur les résultats personnalisés). Mais ce mécanisme, pour diverses raisons, reste mineur par rapport à la dialectique : similarité/importance

Le problème des requêtes larges sur un index en expansion !

Avec la croissance du World Wide Web, et de la « profondeur » des index de moteurs (c'est-à-dire leur taille pour simplifier), les requêtes se sont mises à renvoyer de plus en plus de résultats. L'économie réelle a envahi le web, et beaucoup d'univers sont devenus concurrentiels. Aujourd'hui, sur les requêtes marchandes, les 30 premiers résultats (les 3 premières pages) au moins sont occupés par des URL de sites marchands qui optimisent leurs sites pour les moteurs.



Evolution du nombre de « sites web » (hostnames) depuis 1995. La croissance continue du web a connu trois coups d'arrêt : un premier en 2002/2003, un bref recul en 2007, suivi d'un plateau, et un hoquet plus récemment en 2009/2010

En fait, dès que la requête devient concurrentielle, on va assister probablement aux deux phénomènes suivants :

- Les pages étant très nombreuses, les premiers résultats risquent (c'est statistique) d'être, non seulement proches de la requête, mais vraiment proches entre eux (notes quasi identiques) ;
- La redistribution par les notes d'importance des pages joue donc à plein, et sur un nombre considérable de résultats.

Finalement, on se retrouve avec un jeu de résultats caractérisé par :

- Des notes de similarité sémantique avec la requête très proches ;
- Des notes d'importance et de popularité également très proches ;
- Un classement en tête des résultats essentiellement fait sur des critères d'importance et de popularité.

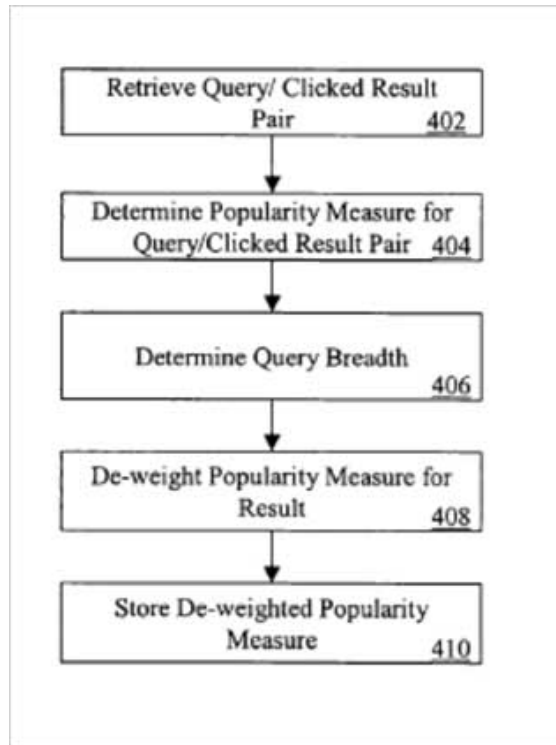
Le problème avec les signaux de type « importance/popularité », c'est qu'ils n'ont rien à voir avec le contenu de la page. Cette situation ne garantit pas que les réponses en première page soient les meilleures, loin de là.

En fait, si l'on en croit les auteurs du brevet déposé par Google, les évaluations ont démontré que ces mécanismes dégradaient la qualité des résultats sur les requêtes larges, et qu'il fallait proposer un mécanisme pour remettre le contenu au centre du classement de l'algorithme.

La solution : redonner du poids aux critères de similarité (on page) ?

La solution proposée par les auteurs du brevet est d'une très grande simplicité. L'algorithme de base du moteur s'appuie sur un équilibre subtil entre les signaux de type proximité/similarité et importance/popularité. Ces derniers doivent avoir un poids suffisant pour départager les ex-aequo, mais pas plus.

Puisque, par un phénomène de sélection statistique, les critères d'importance et de popularité des pages prennent un poids plus élevé que souhaitable sur les requêtes larges, il convient de diminuer le poids des critères de popularité sur les requêtes larges, pour redonner plus de chances aux pages qui ont le meilleur contenu.



Le mécanisme de stockage de l'information recueillie sur les requêtes larges décrit dans le brevet déposé par Google

La méthode décrite dans le brevet est-elle réellement utilisée par Google ? On ne peut en être certain. Mais le problème que ce brevet cherche à résoudre existe bel et bien, et on peut au moins supposer qu'un mécanisme de compensation similaire doit exister dans les algorithmes des principaux moteurs.

Il s'agit d'un nouvel indice de l'importance de la catégorisation des requêtes dans les algorithmes récents des moteurs majeurs : tout se passe comme si l'algorithme changeait en fonction des requêtes tapées.

Quelles conséquences pour le référencement ?

Les faits évoqués dans cet article replacent le contenu de la page (les critères « on page ») un peu plus haut dans la hiérarchie des éléments importants, y compris et *a fortiori* sur les requêtes génériques ou larges, même si elles sont ultraconcurrentielles.

Nous avons d'ailleurs été longtemps surpris par les analyses de corrélation entre critères et classements, faites par un site anglo saxon bien connu (SEOMoz pour ne pas le citer) : la place faite aux critères « on page » (au moins la façon dont les résultats étaient présentées) était en fort décalage avec la réalité. Ils se sont d'ailleurs récemment ravisés, après avoir

découvert la (relativement) forte corrélation entre les classements et une mesure de similarité sémantique tirée de la méthode LDA (*Latent Dirichlet Allocation* : un outil qui fera l'objet d'un prochain article certainement).

La course au Pagerank n'apporte donc pas une garantie de présence sur la première page de résultats, et encore moins dans les trois premiers résultats. C'est vrai surtout si le contenu de la page n'a pas la qualité suffisante pour être considérée comme une réponse pertinente à la requête tapée.

D'une manière générale, ce point démontre que l'algorithme de Google est fait d'un amoncellement de signaux, de filtres et de mécanismes complexes, et que la bonne approche pour optimiser ses pages dépend de plus en plus du contexte.

Faut-il inventer un référencement pour chaque type de requête ? C'est sans doute plus complexe, mais c'est une approche qui commence à prendre du sens, en particulier sur des requêtes locales, d'actualité, des recherches sur des noms de personnalités etc.

La compétition étant par ailleurs féroce sur les requêtes larges, il y'a fort à parier que les nombreux changements d'algorithmes intervenus sur ces requêtes en annoncent d'autres : améliorer leurs résultats n'est pas le seul moteur du changement dans les algorithmes, s'adapter aux évolutions du web en est un autre, tout aussi important.

Bibliographie

Brevets

Methods and systems for adjusting a scoring measure based on query breadth

Invented by Karl Pflieger and Brian Larson

<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=/netahtml/PTO/search-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7,925,657.PN.&OS=pn/7,925,657&RS=PN/7,925,657>

Assigned to Google

US Patent 7,925,657

Granted April 12, 2011

Filed: March 17, 2004

System and method for providing search results with configurable scoring formula

http://www.google.com/patents/about/7505961_System_and_method_for_providing.html?id=Pwe6AAAAEBAJ

Inventors: Finger, II; James Charles (Kirkland, WA)

Assignee: Microsoft Corporation (Redmond, WA)

Appl. No.: 10/155,769

Filed: May 24, 2002

Query revision using known highly-ranked queries

[http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1="20060224554".PGNR.&OS=DN/20060224554&RS=DN/20060224554](http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=)

Invented by David R. Bailey, Alexis J. Battle, David Ariel Cohn, Barbara Engelhardt, and P. Pandurang Nayak

US Patent Application 20060224554

Published October 5, 2006

Filed: November 22, 2005

Articles

Hilltop: A Search Engine based on Expert Documents

<ftp://ftp.db.toronto.edu/pub/reports/csrg/405/hilltop.html>

Krishna Bharat

Compaq, Systems Research Center, Palo Alto, CA 94301

George A. Mihaila

Department of Computer Science
University of Toronto.

The Structure of Broad Topics on the Web

WWW '02 Proceedings of the 11th international conference on World Wide Web

<http://www.cse.iitb.ac.in/soumen/doc/www2002t/p338-chakrabarti.pdf>

Soumen Chakrabarti, Mukul M. Joshi, Kunal Punera, David M. Pennock

Query formulation in web information search

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.76.8657&rep=rep1&type=pdf>

Anne Aula

Tampere Unit for Computer-Human Interaction

Department of Computer and Information Sciences

Pinninkatu 53B, FIN-33014 University of Tampere

To Personalize or Not to Personalize: Modeling Queries with Variation in User Intent

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.146.5090&rep=rep1&type=pdf>

Jaime Teevan, Susan T. Dumais, Daniel J. Liebling

Microsoft Research Redmond, WA 98053 USA

Philippe YONNET, *Directeur SEO international, Twenga.*

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

<http://blog-abonnes.abondance.com/2011/07/lalgorithme-de-google-est-il-modifie.html>