

La gestion des requêtes utilisateur par les moteurs de recherche

[Retour au sommaire de la lettre](#)

| | | |
|------------------|-----------|----------------------|
| Domaine : | Recherche | Référencement |
| Niveau : | Pour tous | Avancé |

On le sait, les moteurs de recherche stockent pendant une certaine durée les requêtes saisies par les internautes sur leurs outils, en plus de certaines informations comme l'adresse IP, etc. L'un des prétextes donnés par ces moteurs pour stocker ce type de données est que cela les aide à bâtir de meilleurs moteurs et algorithmes de pertinence. Mais à quoi peuvent bien servir ces informations ? Comment sont-elles traitées par les moteurs et en quoi cela les aide-t-il à être meilleurs au quotidien ? A la réflexion, on peut trouver de très nombreuses voies d'amélioration des outils de recherche basées sur ces données. Démonstration...

Lorsqu'on s'appelle Google, Yahoo ou Bing, et que l'on reçoit chaque jour un nombre faramineux de visites d'utilisateurs tapant un nombre encore plus grand de requêtes, on dispose en pratique d'une source de données formidable pour comprendre le comportement de recherche de ses utilisateurs. Les moteurs de recherche enregistrent en effet dans un fichier *ad hoc* (un fichier de "log") l'historique des requêtes tapées afin de l'utiliser pour effectuer ensuite des analyses. Google, en particulier, exploite visiblement de manière intensive ces données...

Nous allons voir que les données issues des logs de requêtes peuvent être utilisés pour des usages divers, et parfois inattendus.

A quoi ressemblent les données stockés ?

Les logs de requête gérées par les moteurs stockent généralement l'IP de l'utilisateur, les mots clés tapés, et un *timestamp* (l'horodatage de la requête). Dans le même fichier de log, on ajoute aussi parfois le tracking des clics de l'internaute sur la page de résultat, comme dans cet exemple issus des logs du moteur Excite (l'adresse IP a été enlevée pour préserver l'anonymat de l'utilisateur).

```
[10/09 06:39:25] Query: holiday decorations [1-10]
[10/09 06:39:35] Query: [web]holiday decorations [11-20]
[10/09 06:39:54] Query: [web]holiday decorations [21-30]
[10/09 06:39:59] Click: [webresult][q=holiday decorations][21]
http://www.stretcher.com/stories/99/991129b.cfm
[10/09 06:40:45] Query: [web]halloween decorations [1-10]
[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
[10/09 06:52:18] Click: [webresult][q=home made halloween decorations][8]
http://www.rpmwebworx.com/halloweenhouse/index.html
[10/09 06:53:01] Query: [web]home made halloween decorations [11-20]
[10/09 06:53:30] Click: [webresult][q=home made halloween decorations][20]
http://www.halloween-magazine.com/
```

La session de recherche de l'utilisateur est ainsi facile à retracer : après avoir tapé la requête "holiday decorations", l'utilisateur ne trouve rien qui l'intéresse sur la première page de résultats. Il clique sur le lien vers la deuxième page de résultats, puis la troisième. Il clique sur le premier lien de la troisième page avant de décider de reformuler sa requête en "home made halloween decorations".

Bien sûr, il est possible également d'ajouter dans les logs l'identification du cookie du navigateur pour avoir des données plus précises encore.



Tag cloud réalisé à partir des requêtes les plus tapées dans le moteur AOL en 2006

Le problème de l'anonymisation des données

Très tôt dans l'histoire des moteurs de recherche, le stockage de l'historique des requêtes est apparu comme un véritable problème en matière de respect de la vie privée des utilisateurs. En effet, il suffit de repérer les requêtes contenant des mots clés jugés "déviant", "illégaux", ou plus prosaïquement liés à l'orientation sexuelle par exemple, pour remonter jusqu'à l'utilisateur grâce à son adresse IP. Mais ces informations permettent aussi de "profiler" efficacement les utilisateurs en fonction de leurs goûts, de leur comportement d'achat, de leurs sujets d'intérêt etc. Entre nécessités de justice et de publicité, tout est dans la bonne mesure...

Parfois contraints par la loi ou les organismes de protection de la vie privée, la plupart des moteurs de recherche ont donc choisi d'"anonymiser" les données contenues dans ces logs, de façon à éviter que l'on remonte jusqu'à l'utilisateur (ou du moins sa machine ou son entreprise). La solution consiste à utiliser un identifiant déconnecté de l'adresse IP. Voici par exemple la structure type des données de logs de requête utilisé par AOL jusqu'en 2006 :

<AnonID, Query, QueryTime, ItemRank, ClickURL>

AnonID, comme son nom l'indique, est l'identifiant "anonymisé" de l'utilisateur (enfin du couple navigateur + IP pour être plus précis).

En réalité, cette anonymisation n'est pas parfaite. Pour des raisons légales, les moteurs conservent des données brutes dans leurs logs serveurs, il est donc possible de remonter jusqu'à l'utilisateur malgré tout. Cela signifie qu'une autorité politique ou judiciaire peut se voir donner accès à l'historique des requêtes, ou qu'un employé de Google peut remonter jusqu'à vous.

AOL a appris à ses dépens que l'anonymisation des données dans les logs de requête ne suffit pas ! En effet, l'analyse du contenu des requêtes suffit parfois. En Août 2006, pour faire avancer la recherche sur l'*Information Retrieval*, AOL avait rendu public ses logs anonymisés. Très vite, des petits malins se sont rendu compte que l'on pouvait identifier les utilisateurs grâce à leurs requêtes, ce qui a amené AOL à retirer ces fichiers, et à licencier le responsable de la diffusion des données.

Parmi ces petits malins figuraient des journalistes du New York Times qui grâce aux requêtes de l'*AnonID 4417749* ont identifié cet utilisateur comme correspondant à Mme Thelma Arnold, veuve de 62 ans vivant à Lilburn...

Exemple des requêtes qui ont permis d'identifier l'utilisateur 4417749 :

#4417749:
clothes for age 60
60 single men
best retirement city
jarrett arnold
jack t. arnold
jaylene and jarrett arnold
gwinnett county yellow pages
rescue of older dog
movies for dogs
sinus infection
landscapers in Lilburn, GA
....

Les logs des requêtes pour améliorer le fonctionnement du cache

Le temps de réponse est un enjeu très important dans l'élaboration d'un moteur de recherche efficace. Le fait que le requêteur soit capable de renvoyer un résultat à l'utilisateur en quelques dizaines de millisecondes joue beaucoup dans le succès de Google, de l'aveu même de ses fondateurs. Une telle rapidité à fournir des résultats nécessite d'optimiser toutes les étapes du processus, depuis le hardware, le fonctionnement des réseaux, le système de fichiers (le fameux "*Google File System*"), en passant par le système d'accès aux données (*BigTable*), etc.

Les moteurs de recherche s'appuient aussi sur des systèmes de cache performants : quand une requête est tapée souvent, elle est placée dans un cache pour éviter de recalculer la page inutilement. Ce système de cache simple est souvent doublé par un tampon "dynamique" qui stocke temporairement les pages de résultats pour les ressortir si l'on retape la même requête dans les secondes qui suivent.

On peut également stocker en cache les "*posting lists*", c'est-à-dire les données stockées dans l'index pour une entrée (soit pour la requête "[assurance automobile](#)" mettre en cache les données de l'entrée "[assurance](#)" et celles de l'entrée "[automobile](#)"). Les caches des moteurs ont en effet évolué vers des systèmes plus complexes à plusieurs niveaux.

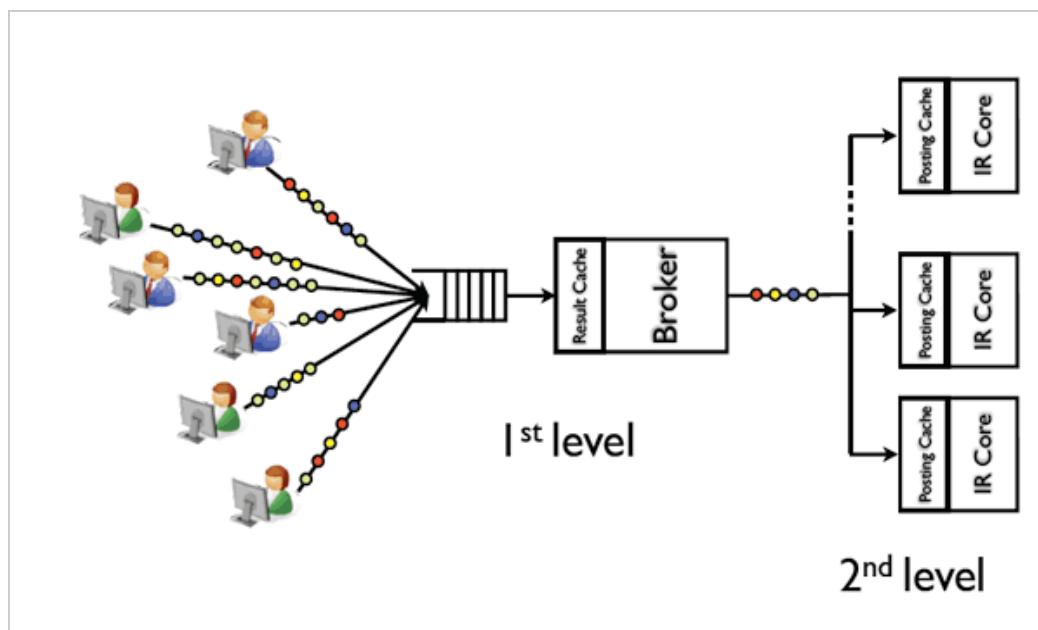


Schéma d'un système de cache à deux niveaux décrit par Ricardo Baeza Yates. Le cache de résultat est placé en amont du "broker" qui assure la répartition des requêtes aux machines ou grappes de serveurs constituant les "IR Cores". Les postings sont mis en cache au niveau de chaque IR Core pour améliorer l'efficacité de l'ensemble.

Tous ces systèmes de cache font également une utilisation intensive des logs de requêtes pour élaborer les stratégies optimales de conservation des données en cache.

Améliorer l'architecture de l'index

A partir de l'analyse des logs de requêtes, on peut travailler sur les fréquences de co-occurrence des termes figurant dans les requêtes. Si deux termes sont souvent associés dans les requêtes, il est intéressant de stocker les données liées à ces entrées dans des zones proches, et même - si possible - sur les mêmes serveurs pour améliorer les temps de réponse.

Extraire des relations sémantiques des logs de requêtes

Mais l'analyse des termes figurant dans les requêtes est aussi une source de données très utile pour identifier des associations de termes. L'analyse des fréquences de co-occurrence combiné avec l'examen des scénarios de requêtes successives permet notamment d'identifier plusieurs types d'association de mots, notamment :

- les expressions "toutes faites", par exemple : "[cul-de-sac](#)" ou "[cheval de Troie](#)" (les mots sont toujours tapés dans le même ordre) ;
- les séquences de mots recouvrant un concept unique ou une "entité nommée" : par exemple "[Président de la République Française](#)" (les mots, là aussi, sont tapés dans le même ordre) ;
- des combinaisons de mots synonymes : puisque les synonymes sont tapés successivement par le même utilisateur au cours de la même session de requête, ces mots sont probablement interchangeables en tant qu'attribut. Ex : [assurance véhicule](#) puis [assurance automobile](#) => *véhicule* et *automobile* sont reliés sémantiquement.

Faire apparaître des suggestions de requêtes

La plupart des moteurs proposent maintenant des suggestions de requêtes qui s'affichent sur la page de résultats. Ces suggestions portent des noms différents selon les moteurs, mais

toutes ces fonctionnalités ont pour but d'éviter à l'internaute qui n'est pas satisfait des résultats d'avoir à retaper une requête complète. Google a attendu longtemps avant de proposer une telle fonctionnalité (en français elle est connue sous le nom : "recherches associées"), elle n'est apparue qu'en janvier 2007 sur un nombre limité de requêtes, avant d'être étendu à une proportion importante des pages de résultats, il y a quelques mois.

Voici ci-dessous un exemple de recherches associées proposées pour la requête "Tour Magne Nîmes". Les termes en gras correspondent à des termes ajoutés ou substitués à ceux saisis par l'internaute. On voit sur cet exemple que les suggestions sont plus ou moins pertinentes, et que les termes associés correspondent soit à des requêtes souvent tapées contenant les termes initiaux de la requête ("[photo Tour Magne Nîmes](#)"), soit des termes associés sémantiquement à la requête : "Nîmes" est une ville du "Gard". Il existe une résidence du CROUS qui s'appelle "Tour Magne" à Nîmes, d'où la proposition "[résidence Tour Magne](#)" et "[CROUS Nîmes](#)".

The screenshot shows a Google search for "tour magne nîmes". The search bar contains the text "tour magne nîmes" and the "Rechercher" button is visible. Below the search bar, there are links for "Rechercher dans : Web", "Pages francophones", and "Pages : France". The search results show "Résultats 1 à 10 sur un total d'environ 34 200 pour tour magne nîmes (0,26 secondes)". On the left, there are links for "Tous les résultats", "Vidéos", "Actualités", "Blogs", "Livres", and "Forums". The main section, "Recherches associées à tour magne nîmes :", displays a grid of suggested search terms:

| | | | |
|--------------------------------------------|---------------------------------------------|----------------------------------------------|-----------------------------------------------|
| résidence tour magne nîmes | résidence tour magne nîmes | jardins de la fontaine nîmes | les celliers tour magne nîmes |
| tour magne nîmes adresse | tour magne nîmes horaires | crous nîmes | hauteur tour magne |
| alla tour magne nîmes | ecole tour magne nîmes | maison carrée nîmes | tour magne arènes de nîmes |
| histoire tour magne nîmes | photo tour magne nîmes | nîmes | tour magne |
| rotary nîmes tour magne | tour magne immobilier nîmes | nîmes | tour magne gard |

Ces suggestions sont construites à partir de l'analyse des logs de requêtes, et d'une logique de combinaison entre ces termes et des termes fréquemment tapés avec ceux de la requête initiale, ou en leur substituant des équivalents et des synonymes. Cette deuxième approche est plus susceptible de produire des résultats de mauvaise qualité. Si l'on tape "Tour Magne à Nîmes" au lieu de "Tour Magne Nîmes", le système suggère de chercher des "Tour Magne" à Alès, Montpellier, Avignon, et "à Gard", ce qui ne risque pas de produire beaucoup de résultats pertinents...

The screenshot shows a Google search for "tour magne à nîmes". The search bar contains the text "tour magne à nîmes" and the "Rechercher" button is visible. Below the search bar, there are links for "Rechercher dans : Web", "Pages francophones", and "Pages : France". The search results show "Résultats 1 à 10 sur un total d'environ 28 000 pour tour magne à nîmes (0,82 secondes)". On the left, there are links for "Tous les résultats", "Vidéos", "Actualités", "Blogs", "Livres", and "Forums". The main section, "Recherches associées à tour magne à nîmes :", displays a grid of suggested search terms:

| | | | |
|------------------------------------------|----------------------------------------------|-----------------------------------------------|------------------------------------------|
| maison tour magne nîmes | résidence tour magne nîmes | à nîmes | tour magne à gard |
| adresse tour magne nîmes | tour magne immobilier nîmes | à nîmes | tour magne à montpellier |
| alla tour magne nîmes | tour magne nîmes france | georges besse à nîmes | tour magne à alès |
| ecole tour magne nîmes | tour carrée nîmes | pont du gard à nîmes | tour magne à avignon |
| photo tour magne nîmes | rotary club nîmes tour magne | jardin de la fontaine à nîmes | tour magne à arles |

A la décharge de Google, la fonctionnalité "recherches associées" ne cesse d'être améliorée et la qualité et la pertinence des suggestions s'améliore de jour en jour.

Les logs de requête servent également à réaliser la fonctionnalité d'autocomplétion des requêtes que l'on retrouve dans beaucoup de moteurs de recherche : par exemple "Google Suggest". A chaque frappe de caractères, le système propose les requêtes les plus probables en fonction de l'historique des requêtes de l'ensemble des internautes. Le système peut, de plus, être personnalisé en fonction de l'historique associé au cookie de l'utilisateur.



Réaliser une expansion de requête

Lorsqu'un internaute tape une requête dans un moteur, il court le risque d'utiliser un terme qui est présent dans certains documents pertinents, mais pas dans tous. S'il tape la requête "assurance automobile", il ne trouvera pas forcément les documents contenant "garantie véhicule". Pour faire apparaître le maximum de documents pertinents, les moteurs de recherche procèdent de manière transparente à une "expansion de requête" en cherchant à la fois les documents pertinents sur la requête initiale et sur des requêtes contenant des mots clés synonymes.

Dans Google, la présence d'une expansion de requête est souvent détectable grâce à la fonctionnalité de "mise en gras" des termes de la requête. Par exemple, si l'on cherche le terme "assurance véhicule" dans Google, on voit que les termes "assurance" et "véhicule" sont mis en gras comme prévu, mais aussi "assurances" et "auto". On voit d'ailleurs que les documents jugés pertinents sur cette requête contiennent les termes "assurance auto" plutôt qu'"assurance véhicule".

Une fonctionnalité similaire dans Google substitue à un acronyme les mots qui servent à le former. Par exemple la requête DIF renvoie aussi des documents contenant les termes "droit individuel à la formation".

Les logs de requêtes combinés au tracking des clics sur les pages de résultats sont intensivement utilisés pour réaliser ce type de fonctionnalités (combiné avec d'autres approches).

Corriger les fautes de frappe et les fautes d'orthographe

Si l'on observe les logs de requêtes, on s'aperçoit que la proportion de requêtes contenant des fautes d'orthographe ou des fautes de frappe est très importante. Les moteurs de recherche cherchent à corriger ces fautes, soit automatiquement lorsque la probabilité d'une faute de frappe est très grande (internet connexion renvoie internet connection en langue anglaise. A noter : dans ce cas, connection n'est pas mis en gras, car il ne fait pas partie de la requête, contrairement au cas des expansions de requêtes décrite précédemment), soit en suggérant la graphie appropriée.

Ci-dessous, on voit un extrait des logs de requêtes concernant britney spears (le nombre à gauche correspond au nombre d'utilisateurs ayant tapé l'orthographe mentionnée à droite). Les graphies différentes sont beaucoup moins fréquentes que la graphie officielle dans les requêtes des internautes, il est donc possible de suggérer la bonne graphie lorsque l'on tape "briteny spears" (image de droite).


```
468941 britney spears
40134 brittany spears
36315 brittney spears
24342 britany spears
7331 britny spears
6633 briteny spears
2696 brittney spears
1807 briney spears
1635 brittney spears
1479 brintey spears
1479 britanny spears
1338 britiny spears
1211 britnet spears
1096 britiney spears
991 britaney spears
991 britnay spears
811 brithney spears
811 brtiney spears
664 birtney spears
664 brintney spears
664 briteney spears
601 bitney spears
601 brinty spears
544 brittaney spears
544 brittnay spears
364 britey spears
364 brittiny spears
329 brtney spears
269 bretney spears
269 britneys spears
244 britne spears
244 brytney spears
220 breatney spears
220 britiany spears
199 britnnev spears
```



The screenshot shows a Google search interface. The search bar contains 'britney spears' and the 'Rechercher' button is visible. Below the search bar, there are radio buttons for 'Web', 'Pages francophones', and 'Pages : France'. The 'Web' option is selected. Below this, there is a link to 'Afficher les options...'. The search results are displayed below, starting with 'Résultats 1 à 10'. The first result is titled 'Essayez avec cette orthographe : [britney spears](#)'. Below this, there is a section titled 'Résultats dans l'Actualité pour [britney spears](#)'. The first article is from 'Plurielles.fr' and is titled 'Britney Spears ne se considère pas comme un sex symbol - Il y a 4 heures'. The article text says: 'Britney Spears confie qu'elle l'étiquette de sex symbol, elle n'y croit pas trop ! La chanteuse n'est pas très à l'aise avec son apparence. ...'. Below the article text, there are two links: 'ACTU - Britney Spears de retour au mois de mars ? - Première - Autres articles (15) »'.

Améliorer l'algorithme grâce à l'utilisation conjointe des logs de requêtes et du tracking des clics

Comme on l'a vu précédemment sur l'exemple des logs Excite, les clics sur les liens proposés dans les SERPs sont trackés également. Qu'ils soient stockés ou non dans le même fichier que les requêtes, 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.

Par ailleurs, l'analyse des scénarios successifs de requêtes et de clics permet d'identifier les requêtes qui appellent une géolocalisation. Avant la mise en place des fonctionnalités de géolocalisation avancée dans les moteurs, qui faussent aujourd'hui l'analyse, les requêtes du type "Hôtel 2 étoiles" étaient forcément accompagnées d'une ville par un utilisateur averti. L'utilisateur inattentif ou inexpérimenté tapait quant à lui d'abord "Hôtel 2 étoiles" puis "Hôtel 2 étoiles Lyon" pour espérer trouver un hôtel. L'analyse des requêtes permet ainsi de conclure que la requête "Hôtel 2 étoiles" requiert sauf exception une géolocalisation.

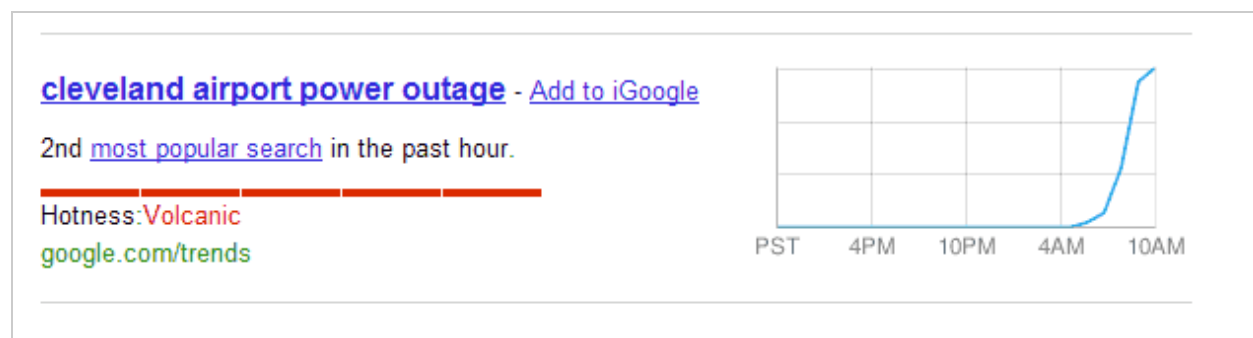
D'une manière générale, l'analyse des scénarios pendant les sessions de recherche permet de "prédire" l'intention de l'internaute qui tape telle ou telle famille de requêtes, et de proposer directement des résultats pertinents.

Personnaliser les résultats

Les logs de requêtes permettent d'associer un historique de recherche à un cookie. Cela signifie qu'il est possible d'utiliser ces données d'historique pour modifier les résultats en fonction de vos préférences passées. Depuis quelques semaines, la personnalisation des résultats, qui ne s'activait que pour les personnes connectées à leur compte Google, s'affiche maintenant pour tout le monde dès lors que le navigateur accepte les cookies de Google.

Analyser la fréquence des requêtes pour détecter les "requêtes qui nécessitent de la fraîcheur"

Si l'on s'intéresse à la fréquence avec laquelle les requêtes sont tapées par les internautes, on s'aperçoit qu'une majorité de requêtes évoluent dans le temps : de nouvelles requêtes apparaissent, d'autres cessent d'être tapées. La "longue traine" réunissant les requêtes qui ne sont tapées qu'à quelques reprises par les internautes est également très importante. Les fonctionnalités Google Zeitgeist, Google Trends, Google Hot Trends ou Google Insights for Search sont des outils qui donnent une bonne idée de cette "volatilité" des requêtes.

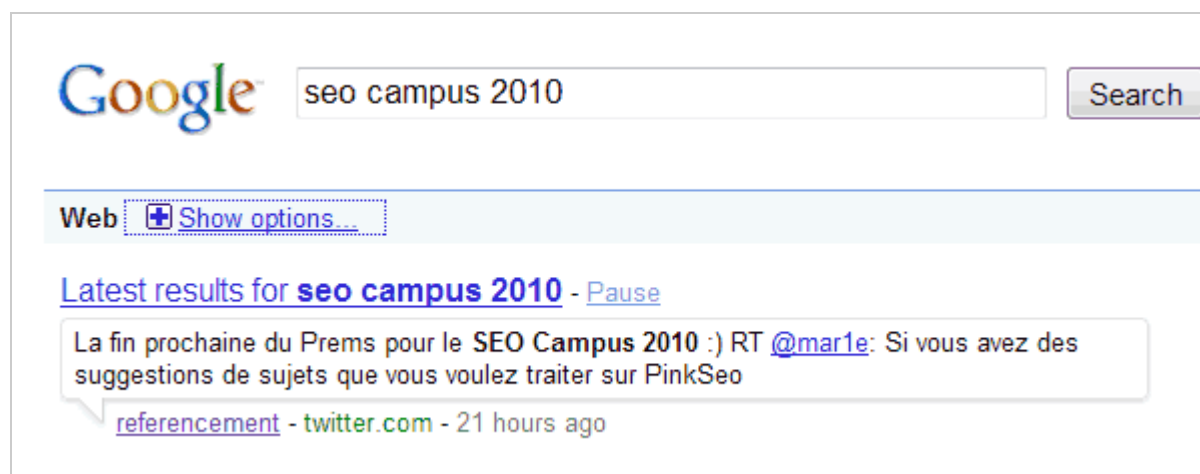


Exemple d'un pic violent de requêtes : une panne d'électricité à l'aéroport de Cleveland bloque le trafic aérien le 10 janvier 2010. En quelques heures la requête devient la deuxième requête la plus populaire sur Google.

Les logs de requête révèlent également que parfois, une requête peu tapée connaît d'un seul coup un pic d'utilisation violent. Pratiquement toujours, ce "pic" révèle qu'un événement s'est produit conduisant un grand nombre d'internautes à taper la même requête en même temps. C'est ce que Google appelle une requête QDF (*Query Deserve Freshness* : la requête nécessite de la fraîcheur). Quand une requête QDF est détectée, Google modifie le classement des résultats pour faire apparaître en priorité des documents récents.

On pourrait penser que la même logique prévaut pour l'affichage de la recherche temps réel dans les résultats. Ce n'est pas tout à fait vrai car une fréquence se mesure sur un intervalle de temps non nul, ce qui peut amener dans certains cas à une réactivité jugée insuffisante. Il semble que la recherche temps réel s'active à partir d'un nombre de requêtes très élevé (le seuil est beaucoup plus élevé que pour une requête QDF) mesuré dans un intervalle de temps très court, ce qui permet par exemple de déclencher en deux minutes la fonctionnalité en cas

de tremblement de terre en Californie (comme le 7 janvier 2010). Enfin, il semble que le contenu des flux temps réel soit analysé pour détecter les événements, et non uniquement les logs de requête de Google. Ce qui explique, par exemple, que la recherche temps réel apparaisse sur la requête SEO Campus 2010, qui apparaissait régulièrement dans les gazouillis Twitter quelques heures avant la rédaction de ces lignes.



Un exemple de résultat de recherche temps réel résultant d'un pic sur Twitter FR

En conclusion...

Les logs de requêtes, combinées au tracking des clics ont toujours été intensivement utilisés par les moteurs de recherche. Au fil du temps, de nouvelles utilisations ont été imaginées. De nouveaux articles ou de nouveaux brevets sont régulièrement publiés, décrivant de nouvelles applications exploitant les données contenues dans ces logs.

Dans bien des cas, leur exploitation suffit à expliquer certains comportements des moteurs attribués à une exploitation de données issues d'autres sources, comme la toolbar de Google.

Avec la généralisation de la personnalisation des résultats (même si elle n'impacte que de manière superficielle les SERPs), ces logs de requête prennent une importance encore plus visible pour l'internaute lambda, puisqu'il voit les résultats "changer" en fonction de l'historique passé de ses recherches.

Mais quelle que soit l'utilité (incontestable) de ces logs pour améliorer le fonctionnement des moteurs de recherche, on ne peut que s'inquiéter de voir qu'autant de données privées - pour ne pas dire intimes - soient en la possession de tiers parfois peu transparents. La phrase d'Eric Schmidt ("*If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place*", prononcée dans une interview de CNBC le 3 décembre 2009) n'est pas pour rassurer les utilisateurs de Google, tant elle semble issue d'un roman du type "1984". Espérons qu'il ne s'agit que d'un dérapage verbal maladroit...

Bibliographie

Applications of Mining Web Queries

Ricardo Baeza-Yates

Yahoo! Research

Barcelona, Spain & Santiago, Chile

http://carbon.videlectures.net/2006/yahoo/fws06/baeza_yates_ricardo/baeza_yates_ricardo_01.pdf

Query-log mining for detecting spam

Carlos Castilloz, Claudio Corsiy, Debora Donatoz, Paolo Ferraginay, Aristides Gionisz

http://airweb.cse.lehigh.edu/2008/submissions/castillo_2008_query_log_detection_spam.pdf

Clique analysis of query log graphs

Alexandre P. Francisco¹, Ricardo Baeza-Yates, and Arlindo L. Oliveira

Clique analysis of query log graphs

<http://research.yahoo.com/files/spire2008.pdf>

The anatomy of a large query graph

Ricardo Baeza-Yates and Alessandro Tiberi

The anatomy of a large query graph

<http://research.yahoo.com/files/JPhys41-2008.pdf>

Query Recommendation using Query Logs in Search Engines

Ricardo Baeza-Yates, Carlos Hurtado, and Marcelo Mendoza

<http://www.dcc.uchile.cl/~churtado/clustwebLNCS.pdf>

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski & Larry Brilliant

research.google.com/archive/papers/detecting-influenza-epidemics.pdf

Query Logs Alone are not Enough

Carrie Grimes Diane Tang Daniel M. Russell

http://www2007.org/workshops/paper_51.pdf

Translating Queries into Snippets for Improved Query Expansion

Stefan Riezler and Yi Liu and Alexander Vasserman

<http://www.stefaniezler.com/PAPERS/COLING08.pdf>

Brevets

Extracting Semantic Relations from Query Logs

Invented by Ricardo Baeza-Yates and Alessandro Tiberi

Assigned to Yahoo

US Patent Application 20090164895

Published June 25, 2009

Filed: December 19, 2007

<http://appft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnethtml%2FPTO%2Fsearch-adv.html&r=1&p=1&f=G&l=50&d=PG01&S1=20090164895.PG.NR.&OS=dn/20090164895&RS=DN/20090164895>

Rank-adjusted content items

Invented by Mayur Datar, Kedar Dhamdhere, and Ashutosh Garg

Assigned to Google

US Patent 7,610,282

Granted October 27, 2009

Filed March 30, 2007

<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnethtml%2FPTO%2Fsearch-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7,610,282.PN.&OS=pn/7,610,282&RS=PN/7,610,282>

Identifying Regional Sensitive Queries in Web Search

Invented by Ya Zhang, Srinivas Vadrevu, Belle Tseng, Gordon Guo-Zheng Sun, and Xin Li

Assigned to Yahoo!

US Patent Application 20090307198

Published December 10, 2009

Filed: June 10, 2008

<http://appft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnethtml%2FPTO%2Fsearch-adv.html&r=1&p=1&f=G&l=50&d=PG01&S1=20090307198.PG.NR.&OS=dn/20090307198&RS=DN/20090307198>

ABBREVIATION HANDLING IN WEB SEARCH

Invented by Xing Wei, Fuchun Peng, and Benoit Dumoulin

Assigned to Yahoo

US Patent Application 20090259629

Published October 15, 2009

Filed April 15, 2008

<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahtml%2FPTO%2Fsearch-adv.html&r=1&p=1&f=G&l=50&d=PG01&S1=20090259629.PGNR.&OS=dn/20090259629&RS=DN/20090259629>

Determining query term synonyms within query context

Invented by John Lamping and Steven Baker

Assigned to Google

US Patent 7,636,714

Granted December 22, 2009

Filed: March 31, 2005

<http://patft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahtml%2FPTO%2Fsearch-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7,636,714.PN.&OS=pn/7,636,714&RS=PN/7,636,714>

Extracting Query Intent from Query Logs

Invented by Priyank S. Garg, Kostas Tsioutsoulis, Bruce T. Smith, and Timothy M. Converse

Assigned to Yahoo!

US Patent Application 20090043749

Published February 12, 2009

Filed August 6, 2007

<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahtml%2FPTO%2Fsearch-adv.html&r=1&p=1&f=G&l=50&d=PG01&S1=20090043749.PGNR.&OS=dn/20090043749&RS=DN/20090043749>

Estimating the Date Relevance of a Query from Query Logs

Invented by Farzin Maghoul and Kostas Tsioutsoulis

Assigned to Yahoo!

US Patent Application 20090043748

Published February 12, 2009

Filed: August 6, 2007

<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahtml%2FPTO%2Fsearch-adv.html&r=1&p=1&f=G&l=50&d=PG01&S1=20090043748.PGNR.&OS=dn/20090043748&RS=DN/20090043748>

Keyword Usage Score Based upon Frequency Impulse and Frequency Weight

Invented by Hua-Jun Zeng, Hua Li, Jian Hu, Han Peng, Zheng Chen, and Jian Wang

Assigned to Microsoft

US Patent Application 20080301117

Published December 4, 2008

Filed June 1, 2007

<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahtml%2FPTO%2Fsearch-adv.html&r=1&f=G&l=50&d=PG01&p=1&S1=20080301117.PGNR.&OS=dn/20080301117&RS=DN/20080301117>

System and method for determining semantically related terms based on sequences of search queries

Invented by Kevin Bartz, Vijay Murthi, Benjamin Rey, and Shaji Sebastian

Assigned to Yahoo

US Patent Application 20080120072

Published May 22, 2008

Filed: November 16, 2006

<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahtml%2FPTO%2Fsearch-adv.html&r=1&p=1&f=G&l=50&d=PG01&S1=20080120072.PGNR.&OS=dn/20080120072&RS=DN/20080120072>

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

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

<http://abonnes.abondance.com/blogpro/2010/01/la-gestion-des-requetes-utilisateur-par.html>