

Les requêtes 'parameterless' sont-elles l'avenir de la recherche sur mobile ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Depuis de nombreuses années, des moteurs de recherche comme Google ou Bing utilisent des données implicites, "devinées", pour améliorer, contextualiser et personnaliser nos résultats de recherche : localisation géographique, historique de requêtes et de navigation, etc. Mais l'avènement des mobiles bouleverse la donne en fournissant de multiples possibilités apportées par les différents capteurs intégrés dans ces terminaux: mesure du bruit ambiant, de la luminosité, de baromètres, etc. Les moteurs peuvent ainsi disposer d'informations appelées 'parameterless' (paramètres implicites) permettant d'affiner des résultats. Ce que des outils comme Siri, Google Now ou Cortana vont exploiter de plus en plus dans les années qui viennent... Quel sera donc leur impact sur nos futures stratégies de référencement ?

Cela fait maintenant plus de 30 ans que les moteurs de recherche grand public offrent une expérience de recherche à base de **mots clés** tapés dans un champ de recherche... Mais l'un des problèmes que pose ce système depuis lors, c'est l'absence de **contexte** autour de ces mots clés. Le contexte d'une recherche (qui fait la requête ?, quand ?, pourquoi ?, où ?, après quelles autres recherches?, etc.) peut permettre en effet de savoir ce que cherche véritablement l'utilisateur qui a tapé un mot clé.

Mais avec l'avènement des smartphones, un certain nombre d'obstacles à la collecte d'informations sur le contexte de recherche ont été (partiellement) levés, et il est devenu possible d'effectuer ce que l'on appelle dans le jargon des moteurs de recherche des "**context-aware queries**" (des requêtes sensibles au contexte) ou des "**parameterless queries**", des requêtes sans paramètres.

Mais de quoi s'agit-il exactement ? Pourquoi ce concept est-il devenu à la mode maintenant ? Et qu'est-ce que cela peut impliquer pour le référencement ? Nous allons nous attacher dans cet article à répondre à ces questions.

Les "context-aware queries", un vieux cheval de bataille des moteurs de recherche

Dans la bataille pour créer un algorithme capable de renvoyer les résultats les plus pertinents pour une requête donnée, faire appel au contexte de la recherche pour améliorer la qualité des résultats a été l'une des premières idées qui est venue à l'esprit des développeurs.

Les requêtes effectuées à base de mots clés sont **ambigües** : par exemple, si quelqu'un tape "jaguar", que cherche-t'il ? Des informations sur l'animal, ou sur la marque de voiture ?

Si l'on vous pose la question suivante : « si je te dis "jaguar", qu'est-ce que tu peux me dire sur le sujet ? », il est fort possible que vous sachiez exactement quoi répondre, pourvu que vous connaissiez le contexte de la question.



Deux exemples de requêtes avec des mots clés « ambigus » souvent citées dans la littérature sur les moteurs de recherche et les techniques de recherche d'information. Si quelqu'un tape « jaguar », cherche-t-il des informations sur l'animal ou sur un concessionnaire jaguar ? Et s'il tape java, veut-il de l'information sur la danse, l'île d'Indonésie, ou le langage informatique ?

Mais de quel "contexte" parle-t-on ici ? En fait, les informations utiles pour "**désambigüiser**" une requête sont de différents ordres.

Le "contexte" est ainsi constitué de différentes informations sur :

- l'état de l'utilisateur et de ses "attributs" (qui est-il ?, quel âge a-t-il ?, que fait-il ?, etc.) ;
- l'état de l'environnement physique de l'utilisateur (où est-il ?, quel temps fait-il ?, quelle heure est-il ?) ;
- l'état de l'application et du dispositif qui sert à faire la recherche (est-ce une tablette, un téléphone, un ordinateur ? Où est ce dispositif ?, comment est-il orienté ?...) ; - et l'historique des interactions entre l'utilisateur, son dispositif, et le moteur de recherche.

Exploiter les informations de contexte est longtemps resté un challenge

Au départ, les premières tentatives pour utiliser le contexte ont échoué sur plusieurs difficultés.

La première est déjà **d'accéder aux informations de contexte**. Il y a une quinzaine d'années, identifier qui était la personne qui effectuait une requête s'avérait compliqué et limité à un petit nombre d'utilisateurs: au mieux on reconnaissait un navigateur grâce à son cookie, et l'intérêt de naviguer "connecté et authentifié" sur AltaVista ou Google était faible. Or si on ne sait pas associer un historique ou un contexte de recherche à un individu, et si on ne sait pas collecter des informations sur son environnement, les informations de contexte sont soit inexploitable, soit trop pauvres pour aider à comprendre le sens d'une requête.

Qui plus est, une certaine "autocensure" a longtemps régné, qui empêchait les moteurs de recherche d'aller trop loin dans la conservation des historiques et des données utilisateurs. Le "tabou" lié aux données personnelles et au "flicage" des comportements a limité les velléités des moteurs sur le terrain de la **personnalisation et de l'exploitation des données utilisateurs**. Le risque "marketing" de voir les utilisateurs se détourner d'un moteur de recherche trop "inquisiteur" a été jugé comme trop élevé pendant la première partie des années 2000. Par défaut, la plupart des moteurs (Google le premier) ont donc longtemps travaillé sur des sessions et des historiques de requêtes "anonymisés". Mais la montée en charge des outils sociaux a "éduqué" les utilisateurs qui sont dorénavant habitués à confier leurs données personnelles contre l'assurance d'un service, et ce "tabou" a de fait disparu depuis une dizaine d'années déjà.


Les derniers obstacles à dépasser pour permettre l'avènement des requêtes non paramétrées étaient en fait liés à la **puissance de calcul nécessaire** et à la **quantité phénoménale de données à stocker et à retraiter** pour parvenir à des résultats probants.

En effet, pour parvenir à améliorer la qualité d'une réponse en fonction du contexte d'une requête, il faut dans un premier temps stocker les données liées à chaque session de recherche, en y ajoutant les informations de contexte (qui a fait cette requête ?, où était-il ?, qu'a-t-il fait avant ?, qu'a-t-il fait après ?, etc.).


Ensuite, il faut analyser toutes ces données pour en tirer des enseignements, et évaluer, lorsqu'une requête est tapée dans un contexte donné, quelle est la probabilité pour que le concept demandé soit, dans le cas du mot clé anglais "gladiator", plutôt lié à une recherche sur l'histoire de la Rome Antique que sur le film de Ridley Scott.

Motivation


Suppose a user raises a query "gladiator"



History?



People?



Film?

If we know the user raises query "beautiful mind" before "gladiator"

- User is likely to be interested in the film
- User is likely to be searching for the films played by Russell Crowe

Exemple d'exploitation des données de session tiré d'un article d'une équipe de Microsoft Research dans leur article « Context-Aware Query Suggestion by Mining Click-Through and Session Data »

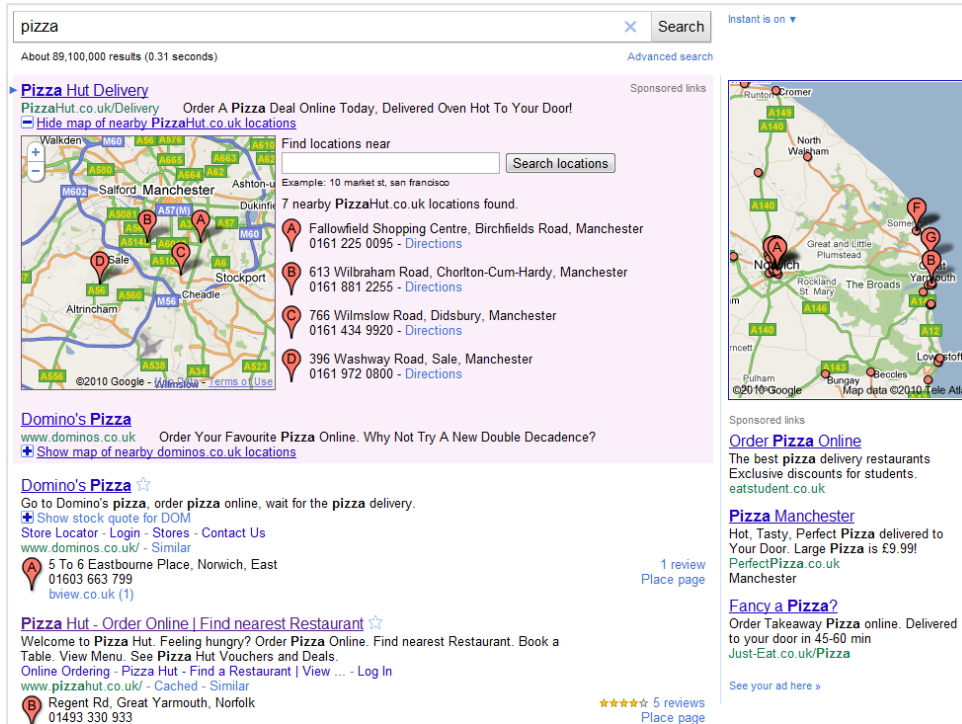
Or le problème est que les données de contexte qui peuvent être stockées sont très nombreuses.

Même si on se limite à celles qui sont porteuses d'un intérêt dans un objectif de désambiguation sémantique, cela augmente malgré tout le volume de données à stocker par le moteur dans des proportions qui ont longtemps découragé les velleités des moteurs de recherche. Car outre le volume de données, calculer comment exploiter ses données de contexte se heurte à une rapide explosion combinatoire liées au nombre de variables à prendre en considération dans l'analyse des comportements utilisateurs.

Les balbutiements des requêtes non paramétrées : les requêtes géolocalisées, ou les requêtes QDF

Même avec peu d'information de contexte, on peut éviter à l'internaute de taper une requête complète, avec tous les éléments (paramètres) permettant de "filtrer" les résultats pertinents et uniquement ceux-là.

Par exemple, dès lors que l'on sait, grâce à son IP, qu'un internaute est basé à Nice, et non à Bordeaux, on peut se servir de cette information pour faire remonter des sites locaux sur la requête "hôtels". Nul besoin que l'internaute précise "Hôtel Nice" pour faire remonter des résultats. Ce comportement de géolocalisation par défaut existe depuis plusieurs années sur Google et Bing.



La requête "Pizza" tapée depuis l'Angleterre renvoie des résultats géolocalisés à proximité de l'internaute. On voit ici aussi les limites d'exploitation du contexte car en réalité le système ne sait pas s'il doit présenter des résultats de Manchester (la préférence de l'internaute enregistrée pour adwords) ou de Norwich (l'endroit où l'IP géolocalise l'internaute de façon erronée).


Le "contexte" peut-être aussi lié à l'environnement de l'internaute. Si une nouvelle fait la une de l'actualité, et si un internaute tape une requête liée à cette actualité, alors le moteur de recherche filtrera en priorité les pages les plus "fraîches" car elles sont probablement plus pertinentes dans ce contexte. Ce type de requête a été affublé du sigle QDF chez Google (*Query Deserves Freshness*).




Google [Recherche avancée](#)
[Préférences](#)


Rechercher dans : Web Pages francophones Pages : France


Web Résultats 1 - 10 sur un total d'environ 83 3

Résultats dans l'Actualité pour **essonne enlèvement**

 **Essonne: enlèvement d'une femme faisant son jogging dans une forêt** - Publié il y a 1 heure
EVRY — Une femme de 42 ans, habitant Milly-la forêt (**Essonne**), a été enlevée lundi vers 09H00 au cours d'un jogging dans un bois de la commune voisine ...
[AFP - 41 autres articles »](#)

ESSONNE : Enlèvement d'une femme partie faire son jogging dans la ...
28 sep 2009 ... **ESSONNE : Enlèvement d'une femme partie faire son jogging dans la forêt** ,
retrouvez l'actualité Société sur Le Point.
[www.lepoint.fr/.../essonne-enlevement.../381191](#) - Publié il y a 1 heure -
[Pages similaires](#) -   

Essonne: enlèvement d'une femme partie faire son jogging dans la ...
28 sep 2009 ... Une femme d'une quarantaine d'années, habitant Milly-la Forêt (**Essonne**), a
été enlevée lundi matin vers 9H00 alors qu'elle faisait son ...
[actu.voila.fr/.../essonne-enlevement-d-une-femme-partie-faire-son-jogging-dans-la-foret_373388.html](#) - Il y a 38 minutes - 

Essonne: enlèvement d'une femme faisant son jogging dans une forêt
28 sep 2009 ... EVRY - Une femme de 42 ans, habitant Milly-la forêt (**Essonne**), a été enlevée
lundi vers 09H00 au cours d'un jogging dans un bois de la ...
[www.lexpress.fr/.../essonne-enlevement-d-une-femme-faisant-son-jogging-dans-une-foret_790802.html](#) - Il y a 15 minutes - 

Exemple de filtrage implicite des requêtes : quelques minutes après l'enlèvement d'une jeune femme dans l'essonne, les 10 premiers résultats renvoyés par la requête « essonne enlèvement » sont des résultats datant de moins d'une heure, sans que l'internaute ait eu à filtrer lui-même sa requête : le paramètre « résultats récents » est implicite.

En fait, les cas dans lesquels les moteurs cherchent à raffiner la requête en ajoutant automatiquement des paramètres tirés du contexte se multiplient. C'est ce que l'on appelle (improprement) des **requêtes sans paramètres**. En fait, les paramètres manquants sont "devinés" par le moteur, et ajoutés automatiquement pour renvoyer du premier coup les résultats pertinents attendus.

Une autre solution traditionnelle pour "raffiner" les résultats consiste à utiliser ce que l'on appelle le "**retour de pertinence**" (*relevance feedback*). Si un internaute tape une requête ambiguë, ou générique, on lui renvoie les résultats de cette première requête, mais aussi d'autres suggestions de requêtes plus "raffinées", que l'utilisateur est invité à faire si besoin est.

La capacité à traiter correctement des requêtes sans paramètres en fonction du contexte est une fonctionnalité plébiscitée par les utilisateurs, car elle est intuitive et adaptée pour des utilisateurs inexpérimentés, ou non concentrés sur la recherche. Les moteurs essaient donc de plus en plus de proposer une approche "parameterless". Les deux solutions sont de toute façon complémentaires.

Les progrès apportés par les smartphones...

Ce qui crée aujourd'hui un intérêt encore accru pour cette approche de requêtes sans paramètres, sensibles au contexte, c'est **l'avènement des usages nomades** en général et des smartphones en particulier.

Déjà, les smartphones sont le plus souvent des objets (très) personnels. Identifier un téléphone, c'est identifier son propriétaire. Là disparaît un des problèmes originels, qui était l'impossibilité pratique de savoir exactement qui se connectait *via* un ordinateur de bureau (potentiellement partagé avec des collègues de travail ou d'autres membres de la famille).

Mais surtout, un **smartphone moderne est un appareil bardé de capteurs**. Sur un appareil haut de gamme en 2014, on trouve pas moins de 11 types de capteurs différents :

- capteurs de lumière et/ou de luminosité ;
- capteurs de proximité ;
- deux caméras (avant et arrière) ;
- plusieurs micros ;
- surfaces sensibles au toucher (écran tactile) ;
- capteurs de position (GPS, Wi-fi, Triangulation Cellulaire, NFC, Bluetooth) ;
- accéléromètre ;
- magnétomètre ;
- hygromètre ;
- baromètre ;
- capteur de température.

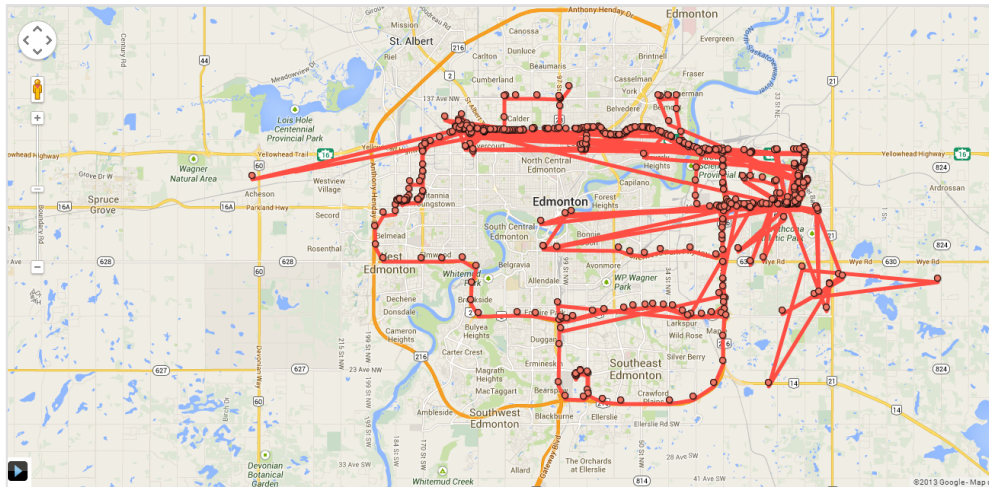
La liste des capteurs embarqués s'allonge progressivement, au fur et à mesure des versions successives des appareils d'Apple, Samsung, Nokia, Sony, etc.



Liste des capteurs innovants embarqués dans les téléphones Samsung Galaxy S4 et S5. On trouve en particulier un gyroscope, un capteur de gestes, une cellule photoélectrique capable de reconnaître la couleur de la lumière ambiante, etc.

Dans ces conditions, certaines informations de contexte, impossibles à réunir auparavant, sont devenues accessibles.

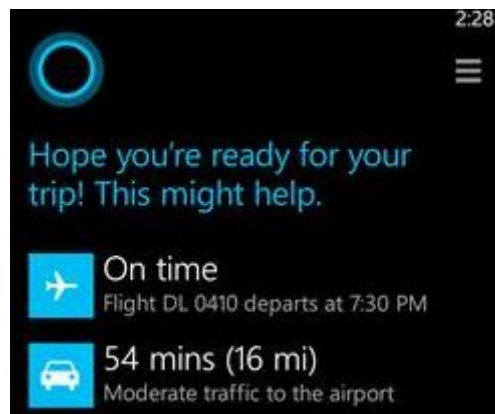
Il est possible de savoir où se situe exactement l'internaute et quel est son environnement (est-ce que l'ambiance sonore est bruyante par exemple). Mais surtout, les informations de navigation et toutes les interactions stockées dans le téléphone permettent d'accéder à des données de contexte très étendues (comme l'historique des endroits visités par le téléphone par exemple).



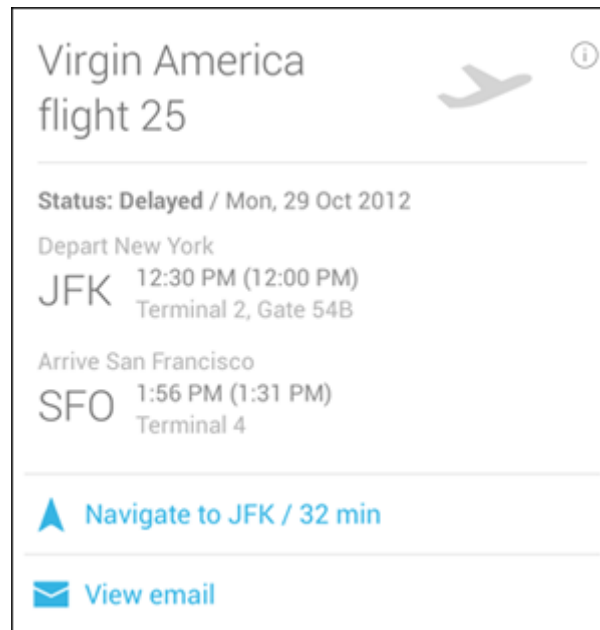
Il est possible de consulter sur Google maps l'historique de tous les déplacements établis grâce aux fonctions de géolocalisation de chacun des dispositifs utilisés par un utilisateur : <https://maps.google.com/locationhistory>

La première génération des requêtes implicites non paramétrées : Google Now et Microsoft Cortana

Lors de leur apparition, les assistants de type "**Google Now**" sur Android ou **Microsoft Cortana** sur Windows Phone ont été remarqués avant tout pour leur capacité à "répondre à la voix" comme Siri, l'outil d'Apple. Mais Cortana et Now aussi intéressants à étudier pour leur capacité à suggérer des informations correspondant à des requêtes implicites (non réellement effectuées par l'utilisateur) et sans paramètres (sans que celui-ci, *a fortiori*, ait précisé quoi chercher, pour quand, et où).



Une suggestion de Microsoft Cortana pour Windows 8.1 : le système sait que l'utilisateur a un avion à prendre, et lui rappelle spontanément son heure de départ et les références de son vol ainsi que le temps de transport jusqu'à l'aéroport en fonction du trafic.



Google Now propose un système similaire

Les "cartes" de Google Now sont en fait des réponses à des **requêtes implicites non paramétrées choisies** en fonction des informations contenues dans le smartphone (comme les rendez-vous mentionnés sur le calendrier, l'historique des déplacements, l'historique des interactions, les photos prises...) ainsi que l'intérêt manifesté par l'utilisateur pour ces "cartes". Les "paramètres" pour chaque requête implicite sont en fait définis grâce aux capteurs et aux informations de contexte fournis par le smartphone (où est l'utilisateur ? où doit-il aller ? quelle heure est-il ? => déclenchement d'une alerte pour lui dire de partir maintenant pour être à l'heure au rendez-vous). Google Now apprend ensuite progressivement quelles sont les requêtes implicites qui semblent utiles pour un utilisateur donné.

Quelle évolution pour Google Now ou Cortana ?

Les versions actuelles de Google Now de Siri ou de Cortana (et leurs concurrents) représentent la première génération de ces outils de recherche sensibles au contexte, et capables de faire des requêtes implicites, et dont les "paramètres" sont tirés du contexte stocké et analysé par le smartphone.

Un brevet déposé par Google fournit un aperçu de ce que pourrait être l'évolution de ce système. L'idée serait de pouvoir "secouer" son téléphone pour lui indiquer que l'on cherche une information. Et ensuite de faire en sorte que l'application de recherche, en se basant sur le contexte du "secouage", devine quelle information va être utile à fournir à l'utilisateur.

Ce système représente en quelque sorte l'aboutissement ultime de l'évolution des requêtes sensibles au contexte et sans paramètres :

- la requête n'est plus du tout formulée : tout est "implicite" ;
- la recherche s'effectue entièrement en fonction du contexte.

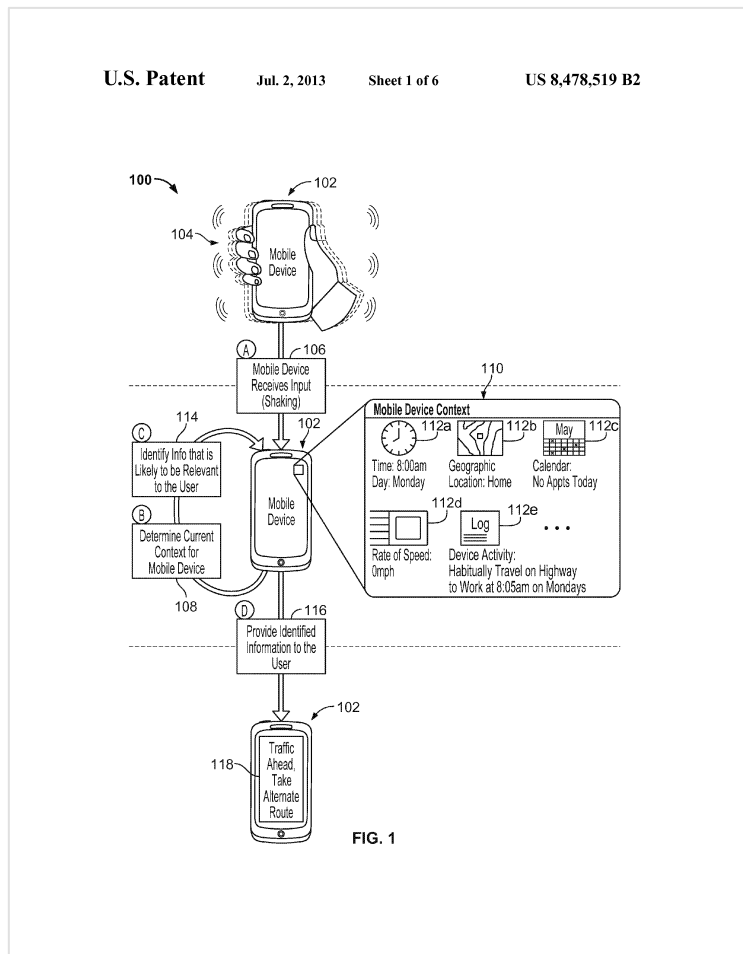


Illustration extraite du brevet déposé par Google « Providing results to parameterless search queries »

Si l'on regarde les informations de contexte que Google propose d'utiliser dans ce brevet, on peut remarquer que toutes ces données sont déjà disponibles sur les smartphones récents et sont en fait déjà utilisées par Google Now. Cette fonctionnalité ne relève donc pas de la Science-Fiction.

Les informations nécessaires seraient :

- la situation géographique du téléphone ;
- les conditions atmosphériques ;
- la liste des entreprises situées dans les environs ;
- le niveau de bruit ambiant ;
- le niveau de luminosité ambiante ;
- les images capturées par la caméra ;
- la vitesse de déplacement du téléphone ;
- la date et l'heure ;
- les rendez-vous du calendrier du smartphone ;
- l'activité récente de l'utilisateur ;
- le comportement habituel de l'utilisateur.

Les requêtes implicites et sans paramètres : quel impact sur les méthodes de référencement ?

Dans l'immédiat, il est peu probable que ce mode de recherche devienne dominant du jour au lendemain. Mais il peut vite devenir populaire et prendre une place de choix à côté des requêtes en langage naturel formulées "à la voix" (Google Voice Search,

Cortana, Siri) et les traditionnelles requêtes par mots clés (avec des paramètres plus ou moins implicites).

Ces requêtes "sans paramètres" constituent un clou de plus dans le cercueil du "paradigme de la recherche par mot clé". Pour fonctionner ainsi, le moteur de recherche doit amasser de plus en plus de "faits" et de "connaissances". Car les résultats de ces requêtes "parameter less" doivent être formulés de la manière la plus directe possible (sinon on perd tout le gain de cette interface de recherche innovante). Dans ces conditions, seules les requêtes navigationnelles risquent à terme de renvoyer vers des pages web dans cette configuration de recherche. Dans tous les autres cas, l'application de recherche va fournir un fait, une donnée, une adresse, un numéro de téléphone, une date... Mais sûrement pas renvoyer vers une page web.

L'enjeu sera donc d'être référencé dans les bonnes bases de données et figurer dans les réponses apportées par ces outils de recherche de seconde génération. C'est donc une toute nouvelle série d'optimisations et de stratégies de référencement et de positionnement qu'il va falloir inventer pour faire partie des "réponses" renvoyées par cette approche nouvelle de la recherche d'information.

Bibliographie

Page d'accueil de Google Now

<http://www.google.com/landing/now/>

Contextual awareness Cortana

<http://www.bing.com/dev/en-us/contextual-awareness>

Comment utiliser Cortana en avant première en dehors des Etats Unis

<http://unleashthephones.com/2014/04/14/how-to-get-cortana-working-on-your-windows-phone-outside-the-usa/>

Brevet sur les requêtes sans paramètres

Providing results to parameterless search queries

Invented by Sumit Agarwal, Vic Gundotra, Alex Nicolaou

Assigned to Google

US Patent 8,478,519

Granted July 2, 2013

Filed: August 30, 2010

<http://patft.uspto.gov/netacgi/nph->

[adv.htm&r=1&f=G&l=50&d=PALL&S1=08478519&OS=PN/08478519&RS=PN/08478519](http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetacgi/nph-adv.htm&r=1&f=G&l=50&d=PALL&S1=08478519&OS=PN/08478519&RS=PN/08478519)

Articles scientifiques

Context-Aware Ranking in Web Search

Biao Xiang, Daxin Jiang, Jian Pei, Xiaohui Sun, Enhong Chen, Hang Li

Microsoft Research Asia

<http://research.microsoft.com/en-us/people/djiang/sigir10.pdf>

Context-Aware Query Suggestion by Mining Click-Through and Session Data

Huanhuan Cao, Daxin Jiang, Jian Pei, Qi He, Zhen Liao, Enhong Chen, Hang Li, KDD'08, 875-883. SIGKDD'08 Best Application Paper Award.

<http://research.microsoft.com/en-us/projects/QuerySuggestion/>

Context-aware Querying for Multimodal Search Engines

Jonas Etzold, Arnaud Brousseau, Paul Grimm, and Thomas Steiner

Erfurt University of Applied Sciences, Germany & Google Germany GmbH, ABC-Str. 19, 20354 Hamburg, Germany,

<http://static.googleusercontent.com/media/research.google.com/fr//pubs/archive/37423.pdf>

Location-Aware Type Ahead Search on Spatial Databases: Semantics and Efficiency
Senjuti Basu Roy and Kaushik Chakrabarti
June 2011

<http://research.microsoft.com/apps/pubs/default.aspx?id=145446>

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