

Interview : Craig Silverstein, Directeur de la technologie (Google)

[Retour au sommaire de la lettre](#)

Une interview de Craig Silverstein, Directeur de la technologies de Google, que nous avons rencontré dernièrement à Paris. Thème de l'entretien : "à quoi ressemblera Google d'ici quelques années ?"

Craig Silverstein, "Director of Technology" de Google (<http://www.google.fr/press/management.html#craig>) était à Paris fin mars. Craig est l'un des pionniers de Google (il a été l'un des premiers employés embauchés par les deux co-fondateurs, Sergey Brin et Larry Page) et connaît parfaitement tous les méandres du moteur de recherche. Nous l'avons rencontré durant une heure et demi. Une bonne occasion pour parler avec lui de Google, de l'avenir des moteurs et de la recherche d'information sur le Web en général.



- Craig, merci d'avoir accepté de répondre à nos questions. Tout d'abord, parlons d'avenir et des grandes voies de réflexion pour améliorer la pertinence des moteurs dans les années à venir. Commençons, peut-être, par la régionalisation, puisque Google vient de sortir le site Google Local (<http://local.google.com/>)...

Oui, en fait, le site Google Local constitue la deuxième étape de notre processus de localisation des recherches. La première étape, débutée depuis de nombreuses années, a été de créer des sites régionaux, pour de nombreux pays à travers le monde, comme [Google France](#) par exemple. Nous croyons beaucoup à ce système de régionalisation. Nous allons maintenant observer et analyser comment les internautes utilisent Google Local, déployer ce type de site dans d'autres pays, pour prendre ensuite des décisions sur la troisième étape du processus que nous ne connaissons pas encore. Nous travaillons beaucoup sur ce point.

- Qu'en est-il de la contextualisation ? Par exemple, le fait d'effectuer des recherches non pas sur la totalité de l'index de Google, mais uniquement sur un "lot" de pages traitant d'une thématique particulière ? En d'autres termes, verrons-nous apparaître un jour des "Google Sports", "Google Santé", etc. ?

Non, ce type de site n'a que très peu de chances de voir le jour chez nous. Nous ne cherchons pas à créer des "Google verticaux". Limiter une recherche à une partie du Web n'est pas dans la philosophie, la logique, de Google. Soit l'utilisateur doit utiliser des mots clés plus précis pour limiter sa recherche à une thématique particulière, soit c'est à nous, sur la base des mots clés demandés, de comprendre la requête proposée, d'être le plus pertinent possible. Nous avons un véritable challenge à relever entre simplicité du moteur et puissance de traitement. Nous avons fait le choix de lancer certains sites, comme Froogle ou Google Local, de façon délocalisée par rapport au site "mère" www.google.com mais la philosophie "idéale" de Google serait de tout rassembler sous une même interface, un seul site. Nous ne devons pas demander à l'internaute de faire le choix d'utiliser tel ou tel outil, c'est à nous de nous adapter en fonction de sa demande. Pour l'instant, les ordinateurs ne sont pas aussi efficaces que les humains pour trier l'information, aussi nous mettons en ligne quelques sites spécifiques, mais réellement le moins possible. Nous ne devons pas nous disperser.

- Qu'en est-il de la personnalisation des recherches ?

La localisation des recherches est l'un des premiers aspects de la personnalisation, qui est l'un de nos grands axes de réflexion pour l'avenir. Comprendre les recherches de l'internaute pour mieux s'adapter à sa demande. Nous travaillons énormément sur la voie de la personnalisation pour améliorer la pertinence de notre moteur, comme pour le lancement du site Google Personal (<http://labs.google.com/personalized>), aux Etats-Unis, qui est, là aussi, une première étape dans notre processus de réflexion. Nous travaillons, bien sûr, sur la compréhension des dernières

requêtes effectuées par l'internaute pour affiner une recherche, mais également sur l'analyse de son adresse IP pour lui proposer des résultats localisés dans sa région, par exemple. La personnalisation est réellement l'une de nos voies majeures de réflexion actuellement.

- La sémantique est-elle dans vos projets également ?

La sémantique est certainement parmi les problématiques les plus complexes du Web et de l'informatique au sens large à l'heure actuelle. Nous n'en tirons la substantifique moëlle que dans 300 ans environ (*rires*). Notre vision est peut-être plus simple et se base sur les commentaires ou les résumés des pages que nous trouvons sur le Web. Le texte des liens pointant vers un document (texte cliquable) en est un bon exemple. Il s'agit d'un champ qui résume souvent bien le contenu d'une page et que nous prenons en compte de façon forte. Pour le reste, il faudra peut-être attendre quelques années pour obtenir des résultats intéressants.

- Tenez-vous compte de la façon dont les internautes utilisent le moteur ? On voit parfois apparaître, dans les résultats, des urls de redirection au lieu des adresses directes des sites proposés. Analysez-vous, par ce biais, le comportement de vos visiteurs ?

Oui, tout à fait. Nous regardons sur quels liens cliquent les visiteurs de Google et, plus globalement, leur utilisation du moteur. Cela nous permet d'analyser et de comprendre certaines choses pour améliorer notre système. Mais ce sont des données collectées uniquement pour notre analyse interne. Il s'agit d'améliorer le moteur par ce biais, rien de plus.

- Les réseaux sociaux comme Orkut font-ils partie de vos projets de R&D ? Peuvent-ils être intégrés à court ou moyen terme au moteur ?

Orkut (<http://www.orkut.com/>) a été développé par l'un de nos développeurs sur ses 20% de travail connexe auquel il a droit chez nous. Pour l'instant, nous regardons l'outil sans réellement savoir si cela peut aider Google à être plus pertinent. Nous allons analyser le projet au fur et à mesure de son avancée et verrons plus tard. Bref, pour l'instant, c'est "wait and see" à ce sujet...

- L'index de Google est aujourd'hui fortement spammé et l'objet de nombreuses attaques de spam et de Google Bombing. N'y a-t-il pas trop de personnes qui savent comment spammer Google ?

Je dissocierais, personnellement, le Google Bombing du spam qui, à notre sens, est plus problématique. Les algorithmes ne sont pas parfaits. Nous travaillons dessus pour que nos résultats soient aussi pertinents que possible. En fait, nous ne luttons pas de façon frontale contre le spam. Nous prenons en compte une politique plus générale. Le spam fait ressortir certains problèmes que nous intégrons dans notre philosophie de pertinence des résultats. Nous ne cherchons pas à combattre ce qui est mal mais à mettre en valeur ce qui est bien...

- Ne faudrait-il pas clarifier, une bonne fois pour toutes, la notion de spam en indiquant noir sur blanc ce qui est interdit.

Ce n'est pas si simple, ce n'est pas notre façon de voir les choses. Notre logique n'est pas d'autoriser ou d'interdire certaines pratiques. L'essentiel est d'aller dans la voie de l'internaute, dans le sens de la qualité du contenu. Interdire n'est pas la bonne façon d'aborder le problème. Si une pratique utilisée sert à mettre en valeur un contenu de qualité, cela peut ne pas poser de problème. Il est donc difficile d'être intransigeant et trop directif dans ce domaine...

- Travaillez-vous sur d'autres voies de réflexion ?

Notre réflexion se base sur trois grandes idées très excitantes à mon avis :

1. L'amélioration continue de la pertinence de nos résultats, notamment en élargissant le champ des documents recherchés, comme avec notre nouveau service, "Google Print", qui propose notamment des résumés de livres.
2. La mobilité. Aujourd'hui, l'internaute utilise Google assis devant son ordinateur. Nous voudrions qu'il puisse le faire également lorsqu'il est mobile, depuis un terminal comme un téléphone portable. Donc, le "wireless" est l'un de nos grands axes de développement. Nous avons ainsi lancé dernièrement, sur nos labos, une version de Froogle qui exploite cette vision.

3. Des voies plus simples pour trouver l'information sous la forme de "raccourcis". Les fonctions "conversion", "calculatrice", "horaires d'avion", etc. en sont un exemple. Il s'agit là d'une façon pour nous de mieux comprendre la requête de l'internaute, au travers de l'interface "classique", et de lui délivrer un résultat le plus proche possible de ses attentes. Bref, le but n'est pas obligatoirement de rendre disponible une info qui n'était pas accessible jusqu'alors, mais plutôt de rendre plus facilement accessible une info déjà disponible...

- Dernièrement, il semblerait que Google ait testé, chez certains utilisateurs, un nouveau design pour ses pages de résultats...

Tout à fait. C'est quelque chose que nous testons régulièrement. Mais cela ne veut pas dire que ces designs seront retenus *in fine*. Il s'agit réellement de tests qui peuvent s'avérer sans lendemain si nous estimons qu'ils ne sont pas concluants (NDLR : *le nouveau look de Google était finalement en ligne quelques jours plus tard...*).

- Vous n'êtes pas sans savoir que Yahoo! a lancé dernièrement sa technologie moteur ;-). N'êtes vous pas un peu choqué que le moteur YST soit un "clone" de Google ?

Google se focalise sur l'internaute, sur l'utilisateur. Nous sommes contents de voir que cela se passe également ainsi ailleurs également. (NDLR : *c'est certainement cela qu'on appelle "une réponse diplomatique" :-)*)

- Pour contrer Yahoo! et bientôt Microsoft, Google va-t-il lancer une nouvelle version révolutionnaire de son moteur ?



Notre philosophie n'est pas de travailler en réaction par rapport à la concurrence. Nous améliorons notre outil de façon continue, quotidienne, sans tenir compte de ce que font les autres. Nous sommes persuadés que la meilleure façon de garder nos utilisateurs, c'est de leur rendre le meilleur service possible. Si nous sommes les meilleurs, les plus pertinents, les plus simples, tout le reste suivra. Ceci dit, l'information est un marché énorme, il y a de la place pour plusieurs acteurs. La concurrence est une bonne chose...

- Si vous ne tenez pas compte de la concurrence, pourquoi avoir annoncé l'augmentation de la taille de votre index, de 3 à 4 milliards de pages, la veille du lancement de YST par Yahoo! ?

Vous devriez peut-être demander à Yahoo! pourquoi ils ont lancé leur technologie juste au lendemain de notre annonce (*rires*)...

- Google va créer des "laboratoires de recherche" à Zurich et à Bangalore. Quels outils seront développés dans ces centres ?

En fait, il ne s'agit pas spécifiquement de "laboratoires de recherche", mais plutôt d'"engineering offices". Les développeurs, dans ces centres, feront le même travail que les personnes situées à Mountain View, en Californie. Ils seront impliqués dans les mêmes univers de projets. Nous avons la volonté d'embaucher les meilleurs informaticiens, mais certains d'entre eux ne désirent pas venir aux Etats-Unis. C'est pourquoi nous avons créé ces bureaux "délocalisés", leur permettant d'être plus proches de chez eux. Il y aura autant de R&D à Zurich et à Bangalore qu'en Californie... Mais pas plus...

- Parlons de référencement payant ("paid inclusion"). Est-il imaginable que Google propose bientôt une telle solution ?

Non. Ce n'est pas notre philosophie, nous l'avons souvent répété. Je ne peux pas croire que Google propose un jour cette fonctionnalité. De même, nous ne proposons pas ce système sur Froogle. Notre logique a toujours été de séparer le gratuit du payant. Nous ne désirons pas intégrer une quelconque offre financière au sein des résultats "organiques" de notre moteur.

- Pourtant, certains webmasters seraient prêts à payer pour avoir une garantie d'indexation dans Google... Ils ne veulent pas disparaître temporairement, pour des raisons techniques, de l'index du moteur...

Il existe plusieurs façons de se prémunir contre de tels désagréments. Par exemple, une meilleure gestion des erreurs 301, au niveau du serveur, avec des redirections bien réalisées. Si le serveur est "down" lors du passage du robot de Google, nous tentons de revenir rapidement et d'analyser au mieux la situation. Nous travaillons sur ces points, mais nous estimons que c'est à nous d'offrir la meilleure qualité de service et que l'éditeur d'un site web n'a pas à payer pour cela.

- Pensez-vous que Google s'oriente vers une mise à jour quotidienne des pages et de son système de PageRank ?

C'est clairement l'un de nos buts. Certaines pages (pas toutes cependant) ont besoin d'une mise à jour quotidienne et c'est notre ambition de fournir à l'internaute une information qui soit la plus "fraîche" possible... Il s'agit d'un axe stratégique prioritaire chez nous. La mise à jour mensuelle est ce qui peut arriver de "pire" sur Google actuellement, sachant que le nombre de mises à jour quotidiennes est de plus en plus important.

- L'augmentation de l'index de Google de 3 à 4 milliards de pages marque-t-elle la limite des systèmes à 32 bits, comme on a pu l'entendre dernièrement ?

Clairement non. D'ailleurs, nous indexons 6 milliards d'objets, si l'on tient compte des documents multimedia. Mais il n'existe pas de limites techniques sur ce point, issus des systèmes à 32 bits.

- Le classement des sites dans les pages de résultats de Google semble évoluer de façon très forte depuis novembre dernier et la Google Dance appelée "Florida". Pourquoi ?

Il faut bien comprendre que nous faisons évoluer notre algorithme de façon continue, quotidienne. Je ne pense pas qu'il y ait eu, depuis novembre, de si grands changements que cela, en tout cas pas plus que d'habitude. Deux facteurs peuvent expliquer cette impression. D'abord, les gens sont peut-être plus attentifs qu'avant au positionnement de leur site sur Google. La période avant Noël a peut-être également exacerbé ce point. D'autre part, nous embauchons beaucoup de nouveaux développeurs et nous sommes donc plus nombreux à travailler sur l'algorithme. D'où des changements peut-être plus importants. Mais, relativement parlant, nous n'avons pas travaillé plus sur l'algorithme depuis le mois de novembre qu'avant cette date.

- Certains disent que Google classent moins bien les sites marchands pour les obliger à prendre des AdWords...

Alors là, je peux vous le dire clairement et dans la langue que vous voulez (*rires*), c'est définitivement NON ! De telles pratiques ne sont pas et ne seront jamais dans la logique de Google !

- Froogle sera-t-il présent en Europe bientôt ?

Oui, Froogle arrivera en Europe car c'est le but de tous nos sites que d'être "régionalisés" à plus ou moins longue échéance. Cependant, pour l'instant, aucune date n'est avancée mais nous y travaillons...

- Enfin, pour terminer, une demande de beaucoup d'utilisateurs avancés de Google : pourquoi la troncature ("tout ce qui commence par...") n'est-elle pas possible sur Google, par exemple à l'aide de l'astérisque, alors qu'elle l'est sur tous les autres moteurs ?

En fait, il s'agit surtout de priorités. L'ajout d'une syntaxe d'interrogation supplémentaire n'est pas, en soi, un projet spécifique chez nous. Nous préférons, en règle générale, porter notre attention aux utilisateurs non avertis en priorité. Bien sûr, nous proposons une syntaxe spécifique et une interface de recherche avancée, mais l'important pour Google reste l'utilisateur "au quotidien". La troncature viendra donc peut-être un jour, il n'y a aucun blocage de notre côté, mais tout est question de temps...

- Merci Craig.

L'entretien s'est déroulé le vendredi 26 mars à Paris, devant un chocolat chaud (pour Craig) et une bière (pour l'intervieweur) :-) Merci à Myriam Boubilil, de Google France, pour son aide. Bonjour à Cindy McCaffrey, vice-présidente de Google, qui était également présente pendant une partie de l'interview...