

Chronique

de la recherche sur le web

par Marc Duval

27 mai 2002, vol.2, n°5

HotBot, l'histoire d'un moteur de la recherche

HotBot a été mis en ligne en mai 1996, soit un an après AltaVista. Il est le fruit de la collaboration entre la société HotWired inc, devenue plus tard Wired Digital. Wired a été ensuite achetée par Lycos en 1999 et en 2000 elle est entrée dans le giron de la société Terra devenant Terra-Lycos. L'origine du nom HotBot n'est pas documentée, mais on peut présumer qu'il vient de la contraction de HotWired's Robot.

HotBot est né de la fréquentation grandissante du site de HotWired et des besoins des utilisateurs. La Société a vite ressenti la tendance de la popularité du web et de la nécessité de répondre aux besoins des internautes devant l'absence d'outil de recherche.

C'est ainsi qu'HotWired s'est associé à l'Université de Californie à Berkeley dont une équipe composée de professeurs et d'étudiants en doctorat travaillait sur une technologie de recherche qui avait pour nom de code Inktomi, nom mythologique donné par les Indiens des Plaines de l'Amérique du Nord à l'Araignée qui a apporté la Culture aux Hommes. A l'époque, HotBot était considéré comme le plus intuitif des sites de recherche, ayant la mise à jour la plus rapide et la plus grande base de données. Il était le concurrent direct d'AltaVista. Les deux sites de recherche ont partagé les honneurs à tour

de rôle jusqu'en 1999 alors que Google émergeait.

Depuis 2 ans, Google est devenu la coqueluche des sites de recherche. Pourtant des liens étroits unissent Google à HotBot. Tout comme Google, HotBot est issu de la recherche universitaire. Il exploite aussi une technologie de calcul parallèle nommée NOW pour Network On Workstation (Réseau de postes de travail - ordinateurs et périphériques) développée par la société Inktomi.

La technologie NOW est basée sur un système informatique à bon marché. C'est un système de réseaux d'ordinateurs adaptable dont on peut augmenter la performance ou la taille de la base de données en ajoutant des composants matériels (machines, disques, mémoires...). Ce système ouvert n'est contraint par aucune amélioration qui pourrait faire en sorte de le remplacer complètement par un matériel plus récent.

Enfin, tout comme Google, il utilise le système d'exploitation Linux.

Mais là s'arrête la comparaison et commence leur différence par les logiciels utilisés.

HotBot utilise les bases de données de Inktomi. En plus de ces bases, Hotbot profite du logiciel Informix d'IBM qui sert entre autres à faire la jonction entre la requête d'un internaute - son profil - et la

publicité, de telle sorte que lorsqu'un internaute fait une recherche sur le thème de l'automobile, le message publicitaire correspondra à ce thème. En fait, Informix est une famille de logiciels de gestion de données dont la fonction est d'analyser ces données, de les synthétiser sous forme de contenu web et de les diffuser.

Le robot de HotBot se nomme Slurp, propriétaire d'Inktomi. [Slurp](#) provient de l'onomatopée du lapement bruyant. En jargon informatique, il pourrait signifier «lire de grande quantité de données avant de les traiter» contrairement à une autre stratégie qui consiste à lire et à traiter des données par petits paquets. Slurp est assisté par le logiciel [SmartCrawl](#) qui a pour fonction de rationaliser la collecte des données afin d'avoir accès à l'information la plus récente. Il peut parcourir jusqu'à 10 millions de documents par jour.

Le site de HotBot se distingue par un formulaire de recherche sur sa page d'accueil et d'une palette de couleurs qui font contraste avec les autres sites qui n'ont qu'une case de recherche et des couleurs sobres. Avec ses formulaires et l'usage que peut en faire l'internaute, HotBot demeure le moteur de la recherche sur le web.

HotBot est ainsi le seul site de recherche qui exploite pleinement les formulaires de recherche bien qu'il ait été doté du langage de recherche le plus évolué. Mais ce langage a quelques

«ratés»: la reconnaissance des majuscules et plus particulièrement la troncature pour HotBot France. En contrepartie, HotBot France est pourvu de la fonction de tri. De plus, HotBot semble avoir une défaillance pour le référencement manuel en regard d'AltaVista, AllTheWeb et Google qui incluent rapidement les URL soumises.

Pour l'instant, sa faiblesse réside à ne pas indexer les documents non HTML. Ce qui ne signifie pas que ces documents ne sont pas accessibles. HotBot a été le premier à offrir la recherche de documents pdf et les images dans les pages web. A vrai dire, Google a fait ombrage à HotBot en affichant ces documents. Jusqu'à tout récemment, HotBot était à égalité avec AllTheWeb. Cette égalité vient de basculer en faveur d'AllTheWeb qui vient d'intégrer les documents pdf à ses résultats de recherche. Il en faudrait peu pour que HotBot rejoigne les prétendants au meilleur site de recherche.

Marc Duval
Service de recherche documentaire DSI
1280-4, bd de Montarville.
Longueuil. Québec. Canada
J4B 8B4
marcduval@dsi-info.ca