

Chronique

de la recherche sur le web

par Marc Duval

12 mars 2001, vol.1, n° 4

Moteurs de recherche ou banques de données web

Qu'est-ce qu'un moteur de recherche? Selon l'Office de la langue française, c'est un « Programme qui indexe le contenu de différentes ressources Internet, et plus particulièrement de sites Web, et qui permet à l'internaute qui utilise un navigateur Web de rechercher de l'information selon différents paramètres, en se servant de mots clés, et d'avoir accès à l'information ainsi trouvée ».

En fait, ce que nous utilisons est une banque de données avec un langage d'interrogation. D'ailleurs, les sociétés telles qu'AltaVista se présentent comme des services de recherche et des index.

Les banques de données web sont alimentées par des webmestres ou des référenceurs et par des logiciels robots parcourant les pages web d'hyperlien en hyperlien. Les services de recherche, sous l'influence de Google, ont mis à profit les hyperliens en utilisant la méthode statistique et relationnel des « liens de notoriété » afin d'améliorer le taux de pertinence des réponses. Ainsi, plus une page contiendra en référence un lien notoire, mieux sera sa position. Cette méthode a cependant des effets... pervers. La société Disney est devenue malgré elle

un site notoire en étant citée sur des pages de site pour adultes comme porte de sortie pour les mineurs.

La mise à jour de ces banques de données varie d'une semaine à plusieurs mois. Une mise à jour trop lente apportera dans les résultats de recherche la présence des erreurs «404» et l'inadéquation entre le sujet de recherche et certaines pages. En raison des délais de ces mises à jour, ces automates de recherche sont inappropriés pour rechercher des nouvelles fraîches. Ainsi, aucun texte de la candidature de M. Bernard Landry à la succession de M. Bouchard n'avait été indexé au 1^{er} février 2001. Par contre, « l'affaire Michaud » (décembre 2000) était présente dans Google.

Ainsi, ces automates de recherche sont-ils excellents pour les archives, notamment Google qui a la particularité de mettre la page indexée à la disposition des internautes.

Marc Duval
Service de recherche documentaire DSI
1280-4, bd de Montarville.
Boucherville, Québec, Canada
J4B 8B4
marc.duval@videotron.ca