

Identificateur formel ou moteur de recherche

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 19 Février 2007

<http://www.bortzmeyer.org/identificateur-vs-moteur-de-recherche.html>

Les deux, évidemment. Pour trouver une ressource sur le Web, les deux sont utiles. Le moteur de recherche permet de trouver des ressources que nous ne connaissons pas et l'identificateur formel permet de désigner une ressource de manière non ambiguë. Alors, pourquoi tant de gens n'utilisent-ils que le moteur de recherche ?

Les utilisateurs dépendent trop des moteurs de recherche et c'est une question sur laquelle les acteurs de l'Internet et les pouvoirs publics, notamment dans l'éducation, peuvent agir. Combien d'utilisateurs ne savent pas utiliser une adresse et dépendent à 100 % d'un moteur de recherche, inconscients du fait que le classement de celui-ci n'est pas et ne peut pas être stable ? Combien de gens savent utiliser <http://www.afnic.fr> (qui est stable) au lieu de taper "afnic" dans Google ? (L'AFNIC a de la chance, elle est classée devant l'Association for Neurologically Impaired Children (<http://www.tsatexas.org/>) mais cela n'est peut-être pas éternel.)

(Je prends Google comme exemple car, en France, la grande majorité des utilisateurs ne connaissent que lui, mais cela serait pareil avec Exalead, Gigablast ou bien Yahoo.)

Dans son livre (<http://www.bortzmeyer.org/google-moi.html>), Barbara Cassin cite le cas d'une pauvre PME qui avait vu son trafic chuter car Google avait modifié ses critères de classement, la reléguant très loin. Mais Google n'a sans doute pas choisi d'embêter cette PME particulière. (Google n'est pas rancunier, taper "Jeanneney" chez eux donne en premier un article sur le directeur de la BNF. Chez Exalead, il n'est qu'en quatrième position.) Non, Google a simplement renvoyé un jour une adresse, un jour une autre, ce qui est la nature d'un moteur de recherche.

Seuls les identificateurs formels comme les URL résolvent ce problème. Un identificateur formel est bien spécifié et a un sens bien précis, évitant ainsi à l'utilisateur de se faire avoir par un intermédiaire qui lui servirait autre chose que ce qu'il a demandé. C'est tout l'intérêt des règles formelles : elles permettent de dire exactement ce que l'on veut dire.

L'intérêt des identificateurs formels est donc leur non-ambiguïté. Sans eux, on ne peut pas désigner une chose particulière, on doit donner des indications vagues et floues (« Tape AFNIC dans Google »), qui peuvent donner le résultat attendu... ou pas.

On lit parfois que la solution serait de remplacer le DNS ou les autres systèmes d'identificateurs formels par un système de mots-clés. C'est une très vieille idée (la mention la plus ancienne que je trouve est dans le RFC 2345¹ mais elle existait certainement avant). Il faudrait se demander pourquoi elle n'a jamais décollé, malgré de nombreuses tentatives commerciales, alors qu'elle est triviale techniquement. (Wikipédia en a une variante, tapez un mot comme « python » dans <http://fr.wikipedia.org/> pour voir.)

Risquons quelques hypothèses pour expliquer l'échec des systèmes par mots-clés, les plus importantes tournant autour de questions non-techniques :

- Il y a toujours des gens qui pensent que **leur** ressource devrait être la seule ou la première (<http://www.bortzmeyer.org/identifiant-convoite-wikipedia.html>) pour un terme donné.
- Cette technique ne résoud en rien les problèmes juridiques. Puis-je acheter le mot-clé « Sarkozy » pour le faire pointer vers <http://sarkostique.over-blog.com/> ?
- Cette technique ne fonctionne qu'avec les protocoles synchrones comme le Web.

Notons bien qu'utiliser les URL et donc DNS n'implique pas de mémoriser des identificateurs comme <http://www-128.ibm.com/developerworks/xml/library/x-matters14.html>. En général, on ne tape pas les URL, on les trouve :

- sur une page Web (imaginez sans le DNS si les liens sur le site Web de l'AFNIC pointaient vers une recherche Google et que "ISC" puisse un jour indiquer l'organisation californienne qui a un rôle important dans Internet et un autre jour l'Independent School Council, qui la frôle dans les réponses de Google).
- par courrier ou messagerie instantanée, ce qui se fait sans rien taper dans presque tous les logiciels modernes.
- ou dans un de ses marque-pages, gardés en local ou hébergés chez, par exemple, del.icio.us.

Dans ces trois cas, rien à taper. Difficile de faire plus simple.

Les utilisateurs sont largement des victimes d'un non-enseignement de l'usage de l'Internet (qui n'est pas l'usage des outils, triviale, mais le bon usage de ceux-ci, ce qui implique de les comprendre). Nous avons là un beau problème d'éducation des utilisateurs, une éducation qui est essentielle mais qui prendra du temps. On constate même des reculs comme l'utilisation de l'interface du moteur de recherche dans les publicités (<http://www.cabel.name/2008/03/japan-urls-are-totally-out.html>), à la place du classique URL.

Un excellent article sur la question est *Qui tient la barre d'adresse?* (<http://thomas-fourdin.net/blog/index.php?post/2007/06/05/Qui-tient-la-barre-dadresse>) de Thomas Fourdin. Citons aussi *Delicious, Chrome et quelques réflexions sur les URI* (<http://www.figoblog.org/node/1928>) d'Emmanuelle Bermès qui fait remarquer (et déplore) que le navigateur de Google, Chrome, renforce délibérément la confusion entre URL et critère de recherche.

1. Pour voir le RFC de numéro NNN, <http://www.ietf.org/rfc/rfcNNN.txt>, par exemple <http://www.ietf.org/rfc/rfc2345.txt>