

Peut-on se passer de moteurs de recherche ?

Stéphane Bortzmeyer
<stephane+crans@bortzmeyer.org>

14 janvier 2012

Exposé libre

Ce document est distribué sous les termes de la GNU Free Documentation License

<http://www.gnu.org/licenses/licenses.html#FDL>.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

Avant, je cherchais, maintenant j'ai Google

<http://normanfaitdesvideos.com/2011/10/19/maintenant-jai-google/>



État de la dépendance

- ▶ Un grand nombre d'utilisateurs ne savent aller **nulle part** sans Google (ou Baidu, s'ils sont chinois),
- ▶ Même à un site qu'ils consultent régulièrement (Facebook, leur banque),
- ▶ En collège, les profs disent aux élèves de B2I « Pour aller sur Wikipédia, tapez "Wikipedia" dans Google ».

Exemple d'incident rigolo

1. Beaucoup d'utilisateurs de Facebook tapent "facebook login" dans Google pour aller à la page d'accueil.
2. Le 10 février 2010, ReadWrite Web publie un article sur l'ergonomie de la page de login de Facebook. RWW est très populaire donc cet article devient le premier résultat de Google.
3. Les utilisateurs de Facebook ayant le plus bas QI laissent des commentaires dans l'article de RWR « *i want to connect to the fafebook plz hlp* » ou « MY HUSBAND WANTS TO SEE HIS ACCOUNT BUT CANNOT LOG IN ggsarla@hotmail.com PLEASE HELP ».

Exemples d'erreurs typiques

[Les exemples sont tirés du journal de mon blog
<http://www.bortzmeyer.org/je-parle-a-mon-moteur-de-recherche.html>]

- ▶ « blog sur les strings » (mots à double sens)
- ▶ « différence entre IANA et AFNIC » ou « revendeur d'objets polonais » (croyance que Google comprend la langue naturelle, l'article comportait bien ces mots mais ne répondait pas à cette question ; peu d'utilisateurs de Google savent qu'il ne connaît pas les **concepts**)
- ▶ « expresso capucino difference » (oubliant que les informaticiens aiment bien les exemples amusants, il est tombé sur un cours SQL, où les données étaient des types de cafés)

Dangers de la dépendance

- ▶ Perte de vie privée : Google sait ce que vous faites et ce que vous cherchez. Si vous tapez « fabrication bombe atomique dans son garage », vous verrez les MIB débarquer... <http://donttrack.us/>
- ▶ Résultats instables : un jour, “Facebook login” mène à Facebook, un autre à ReadWriteWeb... Un jour, “AFNIC” mène au registre de .fr, un autre à une compagnie d’assurances émiratie...
- ▶ Diminution de la résilience : si Google est en panne, l’Internet devient inutilisable,
- ▶ Du fait que tant de gens dépendent exclusivement du moteur de recherche, cela en fait une cible tentante pour ceux qui veulent manipuler les résultats, à des fins commerciales, ou de censure.

Au passage, l’avis des fondateurs de Google...

Sergey Brin et Lawrence Page écrivaient en 2000 dans <http://infolab.stanford.edu/~backrub/google.html>...

For this type of reason and historical experience with other media, we expect that advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers. [...] But we believe the issue of advertising causes enough mixed incentives that it is crucial to have a competitive search engine that is transparent and in the academic realm.

Suite des inconvénients

- ▶ Dépendance via-à-vis de trucs de référencement, au détriment du contenu,
- ▶ Développement du charlatanisme SEO,
- ▶ Polémiques sur le classement (Panda). Un bon classement par Google est-il un Droit de l'Homme ?
- ▶ Plus philosophique : l'adaptation des résultats à votre profil vous coupe-t-elle des autres citoyens, vous enfermant dans une bulle de certitudes? <http://dontbubble.us/>

Mais il y aussi de bonnes raisons

1. Google est **très** rapide (souvent plus que le moteur de recherche du site)
2. Google marche bien (pas d'erreur 500 comme avec tant de moteurs de recherche de site, ou bien des textes non trouvés car comportant des caractères composés)
3. Beaucoup de sites Web sites dépensent plus d'argent à proposer une charte graphique tape-à-l'œil plutôt qu'une navigation compréhensible, *user-oriented*, cohérente et stable. Résultat, le moteur de recherche est nécessaire pour s'y retrouver.
4. Pour certains types de recherche, un moteur est le meilleur outil (par exemple si on ne comprend pas le message d'erreur d'un logiciel, le taper dans Google est en général efficace).

Le problème est que la plupart des utilisateurs se servent du moteur même en l'absence de bonne raison.

Culture

Quelques bons articles et livres sur ce sujet

- ▶ « Google God » d'Ariel Kyrou, un des très rares livres qui ne soient ni *fanboy*, ni vieux con réac
- ▶ <http://thomas-fourdin.net/noms-de-domaine/qui-tient-la-barre-dadresse/122> Une autre conséquence curieuse de la dépendance
- ▶ <http://www.figoblog.org/node/1928> Contre la confusion entre adresse et recherche dans les navigateurs

Les solutions



1. À la dépendance vis-à-vis de Google : autres moteurs de recherche ? Seeks ? Yacy ? Nutch ?

Il y a d'autres techniques de navigation que le moteur de recherche

- ▶ signets (au sens large, pas juste les *bookmarks* du navigateur, cela inclus del.icio.us, SeenThis...)
- ▶ recommandations (je crois qu'on dit curation, pour frimer)

Identificateur formel

Un nom écrit avec une syntaxe rigoureusement définie, et qui identifie un truc (une ressource, une machine, une page...)

Exemple, les URL, les ISBN, les adresses de courrier ou de messagerie instantanée XMPP...

L'identificateur idéal est court, parlant, stable, bon marché, sécurisé... <http://www.bortzmeyer.org/no-free-lunch.html>

Importance des identificateurs

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

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**.

Mais les identificateurs ne sont-ils pas plus difficiles à comprendre ?

1. « La mamie du Cantal ne comprend rien aux URL » (dixit un polytechnicien ou un énarque qui n'y comprend rien non plus)
2. Je réponds : tout est un problème d'éducation. Tout s'apprend. Avec l'imprimerie, il a fallu apprendre à lire et à écrire. Avec l'Internet, il faut apprendre les noms de domaines. Avec le Web, les URL. Un beau chantier.
3. On pourrait commencer par l'école : ne pas apprendre les outils (« cliquez ici ») mais les concepts.
4. Ensuite, apprendre les subtilités des identificateurs qu'on utilise (exemple : les sous-domaines). Cela pourrait aider à la lutte contre le hameçonnage.

Mais les identificateurs ne sont-ils pas plus difficiles à utiliser ?

1. Un coup de main du logiciel peut aider (complétion, recherche DNS en parallèle à la saisie ?),
2. Cliquer n'est pas plus dur que de taper les termes de recherche,
3. Si on veut éviter l'incertitude des moteurs de recherche, il faut de toute façon un petit effort.

Conclusion

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, à la place de l'URL.