



HAL
open science

Google. Une enquête.

Marin Dacos

► **To cite this version:**

| Marin Dacos. Google. Une enquête.. 2005. halshs-00004572

HAL Id: halshs-00004572

<https://shs.hal.science/halshs-00004572>

Preprint submitted on 8 Sep 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

GOOGLE. UNE ENQUÊTE.

Marin Dacos

*Communication prononcée dans le cadre de la journée d'étude « Les enjeux culturels des moteurs de recherche » organisée par la Bibliothèque publique d'information du Centre Georges Pompidou en février 2004. Un débat virtuel a ensuite été ouvert.
<http://debatvirtuel.bpi.fr/moteurs>*

C'est l'histoire d'un responsable de site web, qui met en œuvre une importante refonte technique du site dont il a la charge. Après 18 mois de travail, il inaugure avec fierté la nouvelle mouture de son site. L'ensemble est désormais intégralement dynamique, structuré en XML, les pages sont sémantisées avec le format de métadonnées Dublin Core. Présentation, conformité technologique, simplicité ergonomique, indexabilité. Tout a été prévu pour optimiser le site. Malheureusement, alors qu'il pensait voir la courbe de fréquentation continuer sa progression régulière, il découvre avec horreur que la fréquentation de son site chute brutalement. Il se lance alors dans une enquête qui durera trois mois (au printemps 2003).

Le principal responsable de cette chute est Google. Ce moteur de renommée mondiale, qui n'impose aucun paiement à l'ajout de nouveaux site dans son index, a acquis une position dominante dans les usages des internautes en raison de la rapidité, de la fraîcheur, de la masse et de la pertinence de ses réponses. Mais en l'occurrence, il semble décidé à boudier les nouvelles pages de notre responsable de site web alors que ses anciennes pages, désormais disparues, resteront signalées par le moteur pendant plus de six mois. Cet épisode pourrait suffire à montrer le danger de tout monopole dans la distribution de l'information sur le réseau. Et annoncer des jours sombres pour la liberté d'expression et la circulation mondiale des informations. Mais ici n'est pas le propos. Contrairement à toutes les technologies dites libres, qui sont mises en œuvre en toute transparence technologique, la réussite de Google réside sur un secret. Celui de son algorithme. Bien sûr, tout le monde sait, *grosso modo*, comment le *Page Rank* est calculé. Cet indice sert à classer 4 milliards de pages et à ordonner les résultats en fonction du nombre de liens qui pointent vers chaque page. Cela signifie, clairement, que chaque lien vers un site constitue un « vote ». Plus une page a de votes, plus elle apparaîtra haut placée dans les résultats du moteur. A condition, bien sûr, que Google daigne intégrer la page dans son index.

C'est là que l'édifice commence à se craqueler. Sans entrer dans les détails techniques, il faut raconter la quête du pauvre responsable de site pour réintégrer le gotha mondial... Cela commença par une recherche d'informations. Il découvre alors des centaines de forums discutant exclusivement de la façon dont Google fonctionne. Actifs et nombreux, ces forums sont le lieu où se diffuse, se contredit et se discute la rumeur « Google ». On donne un nom

féminin aux nouvelles indexations de l'Ogre Google, comme on le fait pour un cyclône. On cherche désespérément des informations fiables. On se confronte à un savoir qui semble partiel et ne disposant d'aucune Bible de référence. Notre enquêteur s'adresse donc directement à Google. Il lui faudra patienter mais il obtient des réponses. Toutes plus vagues les unes que les autres. Toutes inadaptées au problème posé. Une sorte de langue de bois technique, accumulant les évidences et les banalités. L'enquête doit reprendre à travers le web. Peu à peu, l'étau se resserre autour d'un forum : www.webmasterworld.com. Dans ce forum dédié aux webmasters, il découvre que des milliers de messages circulent, avec toujours les mêmes questions, avec toujours les mêmes rumeurs, tournant en boucle, à l'infini. Un homme semble attendu comme le Messie. C'est « Google guy ». Cet homme a posté des milliers de réponses aux questions posées sur les forums. Officiellement, sa hiérarchie n'est pas au courant, ce qui justifie l'utilisation d'un pseudonyme. Mais des sites récoltent religieusement les informations qu'il distille ici et là, en constituant de véritables anthologies de la parole de Google Guy. GG, pour les intimes, n'est rien d'autre qu'un outil de propagande financé par Google, qui distille avec succès informations et désinformations en fonction des intérêts de l'entreprise. Tapis derrière l'anonymat et la façade opaque que présente Google, les secrets de l'entreprise qui gère le plus grand et le plus utilisé des moteurs au monde ont de beaux jours devant eux.

Face à tant de stratégies de dissimulation, l'enquêteur décide de passer à l'action. Chaque jour, Google passe sur ses pages. Chaque nuit précédent son passage, il modifie une partie de son site pour voir si la modification permettrait de sortir de l'ostracisme dans lequel il est placé. Chaque lendemain, il constate que ses tentatives se sont soldées par un échec. Jusqu'au jour où une modification, sans doute la cinquantième, se révèle être la bonne. Et en une nuit, ses 20.000 pages sont aspirées par Google. Désormais, il réintègre sa place dans les résultats du moteur. Sa fréquentation récupère en quelques semaines le rang qu'elle occupait avant et elle connaît depuis une croissance continue et forte.

On ne saura jamais pourquoi Google refusait les pages de ce site et la raison pour laquelle il avait organisé tant d'opacité autour de sa façon de procéder. Une rumeur circule, cependant, sur le réseau. Google aurait été dépassé par son succès et par la croissance du web. L'ensemble de son code aurait été calibré pour un maximum de 4 milliards de pages. L'auteur de cette hypothèse pense que Google était alors en train de réécrire l'ensemble de ses logiciels, pour permettre le dépassement de cette limite théorique. Dans cette attente, le nombre de pages du web ne cessant d'augmenter, il aurait fallu trouver une solution pour éviter l'explosion du moteur. Un trop-plein au règles secrètes et au fonctionnement arbitraire aurait alors été mis en place, et des centaines de milliers de pages en auraient fait les frais, en étant exclues de l'index. Fondée ou pas, cette rumeur est à l'heure actuelle la seule explication plausible de l'attitude du moteur.

La conclusion de cette fable, fondée sur des faits réels et qui ne sont malheureusement pas isolés, c'est que la société mondiale mérite que ses principaux fournisseurs d'informations utilisent des règles du jeu claires, explicites et publiques. Or, ce n'est pas le propre des sociétés dont le profit est le principal objectif. Aussi faut-il espérer que naissent des alternatives publiques aux moteurs privés. A l'échelle mondiale, l'UNESCO –par exemple– pourrait se charger de lancer un moteur de recherche généraliste, mondial, de service public, dont les technologies seraient transparentes, les logiciels libres et les résultats gratuits et non sponsorisés.