

**Zeitschrift:** Horizons : le magazine suisse de la recherche scientifique  
**Herausgeber:** Fonds National Suisse de la Recherche Scientifique  
**Band:** - (1998)  
**Heft:** 39

**Artikel:** L'art de chercher  
**Autor:** Giussani, Bruno  
**DOI:** <https://doi.org/10.5169/seals-556165>

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. [Siehe Rechtliche Hinweise.](#)

### **Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. [Voir Informations légales.](#)

### **Terms of use**

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. [See Legal notice.](#)

**Download PDF:** 02.02.2025

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**



Bruno Giussani  
(bruno@giussani.com) est spécialiste  
des nouveaux médias et chroniqueur  
Internet du «New York Times».



# L'art de chercher

**L**es sondages sont unanimes: la difficulté à obtenir rapidement l'information qu'ils cherchent est la principale frustration des utilisateurs de l'Internet.

Les explications avancées sont multiples. Le Web est immense (320 millions de documents selon les estimations les plus récentes) et en croissance rapide. Il est volatile et, contrairement aux bibliothèques et bases de données, l'indexation n'est pas normalisée, chaque site possédant sa propre structure et sa propre façon d'organiser et de présenter l'information.

De ce fait, certains prétendent que la recherche online est un art. Une connaissance minimale de la structure du réseau et du fonctionnement des outils de recherche augmente toutefois sensiblement les chances de localiser rapidement la bonne information. Et une bonne stratégie n'est pas de trop – ce qui commence par une définition claire de ce que l'on cherche: taper «eau» ou «Suisse» ou «Einstein» dans un moteur de recherche ne peut que produire une avalanche de références inutiles ou non pertinentes.

Il existe des centaines d'outils de recherche. D'après leur structure, on peut les classer en deux catégories générales: index ou catalogues thématiques, et moteurs de recherche par mot clé.

Le plus connu des catalogues est Yahoo (www.yahoo.com), qui se présente sous la forme d'un menu thématique. En sélectionnant une catégorie, on accède à une liste de catégories secondaires, et ainsi de suite jus-

qu'à l'information que l'on recherche (si elle est disponible). En utilisant des mots clés, on peut raccourcir la procédure.

Le plus populaire des moteurs de recherche est Altavista (www.altavista.com), dont la base de données est générée par des «robots», ces logiciels qui parcourent le Web, suivent tous les hyperliens qu'ils trouvent, «lisent» tous les documents, et mémorisent la position de chaque mot, l'adresse de la page dans laquelle il se trouve, sa position parmi les autres mots de la même page, et d'autres informations.

Lorsqu'on utilise un de ces systèmes, la recherche n'est pas effectuée directement sur le Web, mais dans la base de données de l'outil en question. Connaître la façon dont ces bases de données sont structurées et apprendre le langage qui permet de les interroger – trouver une phrase particulière, une combinaison complexe de mots, ou circonscrire la recherche à certaines régions du monde, à une période donnée, et même à un site spécifique – et l'utiliser de façon imaginative est donc essentiel.

Par exemple, Yahoo est constitué uniquement du titre, de l'adresse et d'une courte description des pages Web. La probabilité de trouver une page ne dépend donc pas uniquement de sa présence dans l'index, mais également de la façon dont le contenu de la page a été décrit. Le résumé d'une page se référant à la question des «comptes en déshérence» pourrait ne pas contenir le mot «Holocauste» ni «banque»: la référence à cette page ne sera donc pas affichée parmi les résultats d'une recherche sur ces deux termes.

Bien sûr, on ne trouve pas tout sur l'Internet, et même ce qui s'y trouve n'est pas nécessairement indexé par les moteurs de recherche. Malgré leur puissance et sophistication, ceux-ci ne sont déjà plus en mesure de suivre la croissance rapide et désordonnée du Web. Une étude menée début 1998 par des chercheurs de NEC a déterminé qu'aucun moteur de recherche n'indexe plus du tiers du contenu informationnel du Web. Hotbot (www.hotbot.com) en «couvre» 34%, Altavista 28%, Excite (www.excite.com) 14% et Infoseek (www.infoseek.com) 10%.

B.G.