

# Testez la machine à remonter le Web

Les archives du réseau sont désormais ouvertes à tous les internautes.

Los Angeles, correspondance

Depuis leur ouverture au public fin octobre, les plus grandes archives Internet du monde sont prises d'assaut. Alléchés par la perspective de se plonger gratuitement dans une collection de dix milliards de pages web, les visiteurs se pressent mais la moitié d'entre eux est refoulée à l'entrée. Trop d'affluence.

C'est que ce musée en ligne offre une expérience hors du commun: revisiter la brève mais riche histoire du Web. Surnommé «The Wayback Machine» (la machine à remonter le temps), ce site (1) permet de retrouver des pages sauvegardées depuis 1996. Par exemple, le Yahoo! dépouillé des débuts, les premières pages personnelles – touchantes de maladresse et pour beaucoup laissées en jachère – de l'hébergeur gratuit Geocities, les sites (alors) triomphants du commerce électronique, comme l'épicier en ligne Webvan, emportés depuis par le crash de la Net économie. De libération.com, on retrouvera aussi deux numéros datant de 1996, sept de 1997 et jus-

qu'à trois cent soixante de l'année 2000 (plusieurs éditions du même jour ont parfois été archivées).

**Gigantesque.** Mais pour l'heure, ça bouchonne. «On a peut-être été un peu trop pressés d'ouvrir au public. Mais le trafic devrait être maîtrisé d'ici à la mi-novembre», indique le fondateur, Brewster Kahle. Cet entrepreneur, enrichi aux premières heures du Web, a créé,

il y a cinq ans, l'ancêtre de la machine à remonter le temps: «Internet Archive» (2), avec ses propres deniers et une passion sans limites (*Libération*

du 13 octobre 2000). Ses archives n'étaient accessibles jusqu'alors qu'à un groupe restreint de chercheurs et de journalistes. L'an dernier, son collaborateur, Scott Kirkpatrick, reconnaissait ne pas pouvoir imaginer l'ouverture au public avant fin 2002.

Mais les attaques du 11 septembre ont incité Brewster Kahle à généraliser plus tôt que prévu l'accès à ce gigantesque champ de mémoire. Quelque 100000 milliards d'octets d'informations, soit l'équivalent de cinq fois le contenu de la bibliothèque du Congrès à Washington, sauvés de l'oubli. «Si vous ne gardez aucune trace des contenus numériques, ils disparaissent», rappelle ce chercheur retraité d'IBM.

**Embarrassant.** Pour remonter le fil du Web, il suffit d'entrer l'adresse d'un site. «The Wayback Machine» présente alors toutes les versions qui en ont été sauvegardées depuis 1996. Les enregistrements se font au gré de l'humeur vagabonde des «robots» de recherche qui parcourent le Web tous les deux mois, prenant des «photos instantanées» des pages des sites détectés. «Cela amuse les gens de voir ce qui est retenu pour faire partie de l'Histoire», explique Brewster Kahle. Ils retrouvent

parfois des choses embarrassantes comme leur première page web.» Ou des informations gênantes, comme cet extrait du site officiel de la Maison Blanche, «aspiré» il y a cinq ans. Le président Clinton s'engageait alors sur la sécurité dans les aéroports: «J'ordonnerai que la FAA (aviation civile fédérale, ndr) commence à fouiller tous les sacs des passagers des vols intérieurs dans certains aéroports. Et je suis fier d'affirmer que plu-

sieurs recommandations de la commission seront appliquées immédiatement.» Le fondateur des archives a lui-même retrouvé la trace d'un site créé par des amis en Allemagne, disparu dans l'éther du Web avec la faillite de leur hébergeur.

Brewster Kahle rêve finalement d'une bibliothèque d'Alexandrie numérique: «Nous avons la technologie pour créer à moindre frais un énorme fonds, accessible de partout dans le monde.» Numérisé, l'ensemble des ouvrages de la librairie du Congrès tiendrait dans une machine de la taille d'un frigo. Quant aux archives de l'Internet et leurs 100 teraoctets, «ça a l'air gros, mais cela coûte seulement 400000 dollars (442000 euros) à stocker».

**Procès.** L'entreprise fondée par Brewster Kahle, Alexa – désormais filiale du cybermarchand Amazon –, a fourni la technologie nécessaire à l'archivage. Mais le processus est loin d'être parfait. Et de plus en plus complexe. Car depuis cette année, les archives incluent images, sons et vidéos, en plus des textes. «Cela va devenir diffi-

cile à gérer quand tout le monde aura une webcam pointée sur le berceau du bébé», s'inquiète Brewster Kahle.

Mais ce sont plutôt les procès que craint Lawrence Lessig, professeur de droit et penseur du Net. Ce fervent supporter des archives confiait récemment au *New York Times* qu'il ne serait pas surpris que des détenteurs de copyrights se manifestent rapidement. Les pages web peuvent contenir, par exemple, des fichiers MP3 pirates. Mais pour l'heure, Brewster Kahle estime que «90% des gens, dont le site fait partie de la collection, se disent flattés.» Les mécontents peuvent d'ailleurs lui envoyer un e-mail pour demander le retrait de certaines pages.

Chaque mois, une centaine de nouveaux serveurs rejoint le cœur de la machine à remonter le temps, une ancienne baraque militaire, à deux pas du Golden Gate Bridge. Un deuxième local, toujours à San Francisco, devrait compléter le dispositif. C'est que son fondateur voit de plus en plus grand. En sus des collaborations avec les «petits» conservateurs du Web, comme Steve Baldwin et son fameux Ghostsites (3) (*Libération* du 10 août 2001), il espère coopérer avec la France, mais aussi l'Australie et la Suède sur des projets d'archivage de leurs sites nationaux ●

EMMANUELLE RICHARD

- (1) web.archive.org
- (2) www.archive.org
- (3) www.disobey.com/ghostsites