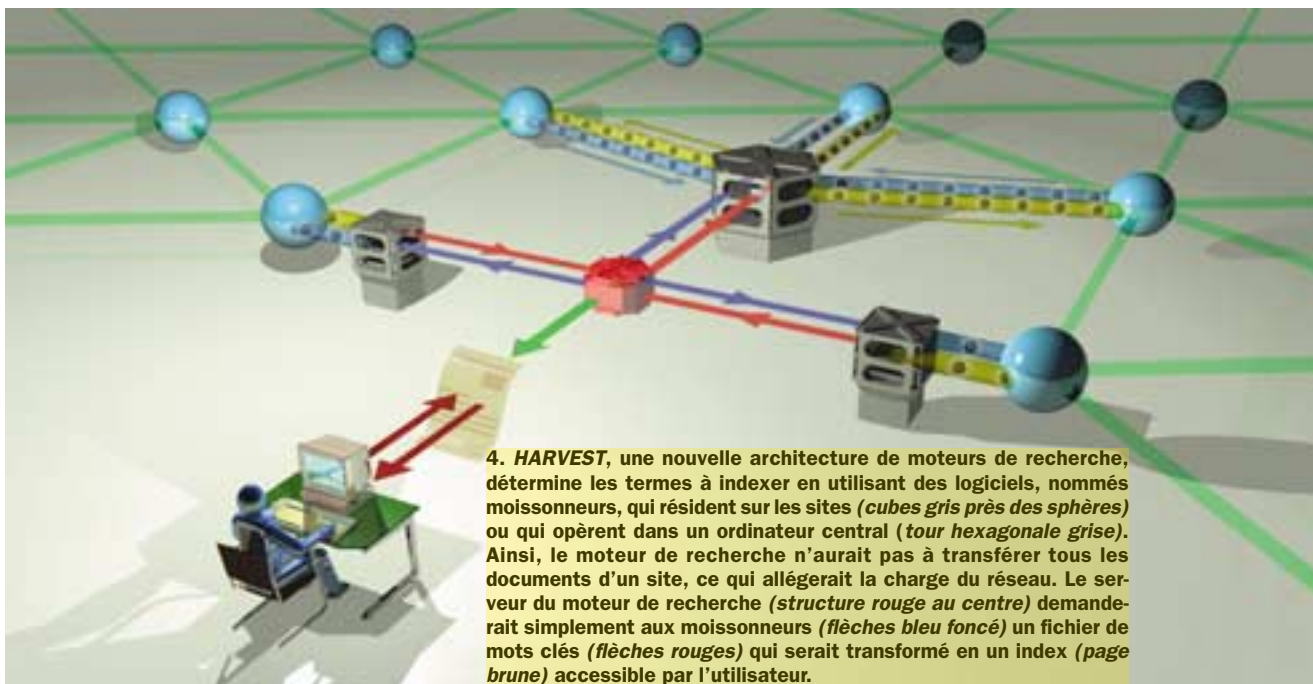




Byron Christie

sujet donné, sans payer les services d'êtres humains chargés de juger la myriade de sites Web. Une solution, qui fait à nouveau appel à l'être humain, consiste à partager entre utilisateurs l'évaluation des sites. Des systèmes d'évaluation numériques récemment mis au point permettent aux utilisateurs de donner leur sentiment sur des sites particuliers (voir *Le filtrage d'information*, par Paul Resnick, page 52).

Certains moteurs explorent l'Internet et, de surcroît, séparent le bon grain de l'ivraie. Il faudrait toutefois de nouveaux programmes pour décongestionner le réseau lors de l'examen répété des sites. Certains gestionnaires de sites affirment en effet que leurs ordinateurs consacrent bien plus de temps à fournir leurs informations aux moteurs qu'aux personnes qu'ils espèrent séduire par leurs offres.

Pour résoudre ce problème, Mike Schwartz et ses collègues de l'Université de Boulder ont écrit un programme, nommé *Harvest*, qui permet à un site de compiler des données d'indexation de ses propres pages et de transmettre, à la demande, ces informations aux différents moteurs de recherche. Ainsi, grâce à *Harvest*, les moteurs n'ont plus à exporter la totalité des sites. Alors que les moteurs de recherche téléchargent un exemplaire de chaque page visitée dans leurs sites d'origine pour en extraire les

termes qui constituent un index, ce qui engorge le réseau, *Harvest*, le «programme moissonneur», n'envoie qu'un fichier de termes indexés. En outre, comme il transmet uniquement l'information concernant les pages qui ont été modifiées depuis sa dernière visite, il évite d'encombrer le réseau et les ordinateurs qui lui sont connectés.

Les programmes moissonneurs pourraient également aider les éditeurs à limiter l'information qui quitte leurs sites. Ce contrôle se révèle nécessaire, car le réseau n'est désormais plus un simple support de libre circulation de l'information ; il devient progressivement un moyen d'accéder, contre paiement, à des informations protégées, dont la lecture n'est pas forcément autorisée pour les utilisateurs. Les moissonneurs, eux, distribueraient uniquement l'information que les éditeurs diffusent gratuitement, tels les sommaires et des extraits de l'information stockés sur leurs sites.

Les besoins des utilisateurs détermineront les méthodes de collecte d'information qui seront mises en œuvre. Il est encore trop tôt pour dire si l'Internet ressemblera alors à une bibliothèque, avec des collections structurées, ou s'il restera anarchique, avec un accès fourni par des systèmes automatisés.

Si les utilisateurs acceptent de payer pour protéger le travail des auteurs, alors les éditeurs, les indexeurs et les critiques privilégiés

ront l'information structurée du type bibliothèque classique. En revanche, si l'information est fournie gratuitement ou subventionnée, l'indexation numérisée, peu coûteuse, l'emportera très probablement, et l'environnement anarchique qui caractérise aujourd'hui une grande partie de l'Internet subsistera. Ainsi, plutôt que des problèmes techniques, ce seront surtout des problèmes économiques et sociaux qui détermineront la forme que prendra, dans l'avenir, la recherche de l'information sur l'Internet.

Clifford LYNCH est informaticien à l'Université de Californie.

C.M. BOWMAN et al., *The Harvest Information Discovery and Access System*, in *Computer Network and ISDN Systems*, vol. 28, n°s 1-2, pp. 119-125, décembre 1995.

Des informations sur le programme *Harvest* sont disponibles à l'adresse Web <http://harvest.transarc.com>

Lorcan DEMPSEY et Stuart L. WEIBEL, *The Warwick Metadata Workshop: A Framework for the Deployment of Resource Description*, in *D-lib Magazine*, juillet-août 1996. Les comptes rendus de cette conférence sont disponibles sur <http://www.dlib.org/dlib/july96/07contents.html>

Carl LAGOZE, *The Warwick Framework: A Container Architecture for Diverse Sets of Metadata*, *ibid.*

Les bibliothèques numériques

MICHAEL LESK

La quantité d'informations disponible sur l'Internet sera très inférieure à celle des futures bibliothèques. Toutefois, la construction de ces géants ne sera pas aisée.

À Paris, au bord de la Seine, quatre tours dominent le quartier de Tolbiac : avec leurs 395 kilomètres de rayonnages, elles pourront recevoir jusqu'à 22 millions de livres. La bibliothèque François-Mitterrand pourrait être la dernière et la première du genre. La dernière, car la plupart des grandes villes n'ont plus les moyens de construire des bâtiments publics aussi ambitieux. La première, car son installation s'achèvera avec le branchement de centaines d'ordinateurs, qui donneront au public un accès électronique immédiat aux textes de 100 000 ouvrages couvrant une grande part de l'histoire et de la culture françaises.

Dans le monde entier, des bibliothèques se sont lancées dans la tâche herculéenne de numériser des livres, des images et des sons, qui forment la mémoire culturelle de l'humanité. Les utilisateurs des bibliothèques devraient bientôt disposer, sans quitter leur bureau, d'une quantité d'informations bien plus considérable que celle qui est aujourd'hui présente sur l'Internet. Cependant, les nombreux obstacles légaux, économiques et techniques rendent cette perspective encore incertaine.

Les bibliothécaires voient trois avantages évidents à la numérisa-

tion. Premièrement, elle leur permettra de préserver des documents fragiles et rares sans en restreindre l'accès. La British Library de Londres, par exemple, détient le seul manuscrit existant du *Lai de Boewulf* (un poème épique anglo-saxon, d'auteur inconnu, écrit entre les VII^e et X^e siècles) ; seuls des spécialistes étaient autorisés à le consulter jusqu'à ce que Kevin Kiernan, de l'Université du Kentucky, ne numérise ce manuscrit sous trois lumières différentes (révélant des détails qui n'apparaissent pas à l'œil nu) et ne mette ces images sur le réseau Internet. De même, la bibliothèque du Parlement japonais est en train de réaliser des clichés numériques, extrêmement détaillés, de 1 236 blocs de bois imprimés, rouleaux et autres matériels qu'elle considère comme des trésors nationaux, afin que les chercheurs puissent les examiner sans manipuler les originaux.

Le deuxième avantage est la commodité : les usagers accèdent aux informations en quelques secondes, au lieu de perdre des heures dans les bibliothèques. Plusieurs personnes peuvent lire simultanément le même livre ou regarder la même image. Les bibliothécaires n'ont plus à remplacer les ouvrages sur les rayons.

Le troisième avantage des copies électroniques est d'occuper un espace de quelques millimètres sur un disque magnétique – au lieu de mètres sur des rayonnages. La construction des bâtiments d'une bibliothèque est de plus en plus coûteuse : la bibliothèque François-Mitterrand coûtera plus de sept milliards de francs ; le stockage de chaque volume reviendra en moyenne à 300 francs (dans des bâtiments moins ambitieux, le coût moyen de stockage varie autour de 150 francs). À titre de comparaison, le prix du stockage sur disque revient à environ dix francs pour une publication de 300 pages et ne cesse de diminuer.

Des compromis techniques

La numérisation sera progressive, car chacune des techniques de numérisation est un compromis entre la conservation, la commodité et le coût. La numérisation des pages en vue d'une création d'images numériques est l'opération la moins coûteuse. Elle revient à environ 200 francs par volume, la majeure partie de cette dépense servant à payer les salaires des personnes qui utilisent le scanner. La technique du *photoCD* de Kodak, qui numérise automatique-



Agence France-Presse



Michel Gauguier, Agence France-Presse

1. LES TOURS de la bibliothèque François-Mitterrand, dans le 13^e arrondissement de Paris (à gauche), ont une capacité de 22 millions de livres. Cependant, 100 000 de ces ouvrages, stockés sous forme d'images numérisées, n'occuperont qu'une place infime et seront instantanément accessibles sur des ordinateurs installés dans les salles de lecture.