

Les interfaces de recherche

MARTI HEARST

Le rapide développement du réseau mondial gêne sa rationalisation et son organisation. De nouvelles interfaces faciliteraient son emploi.

Comment éviter de se perdre dans l'immensité du réseau mondial? Nombre de sites permettent de retrouver facilement certaines informations courantes et sommaires, tels des numéros de téléphone ou les cours de la Bourse, mais le réseau Internet est remarquable parce qu'il ne nous cantonne pas à un quotidien prosaïque et qu'il fait venir sur nos bureaux des informations extraordinairement variées. Pourtant, par manque d'organisation cohérente, le cyberspace devient de plus en plus confus. Les outils qui servent aujourd'hui à localiser un document dans l'Oregon, un catalogue en Grande-Bretagne ou une image au Japon sont souvent d'une lenteur désespérante.

Quelques algorithmes classent les résultats des recherches d'après leur pertinence, mais le remède à la lenteur du réseau résidera plus probablement dans l'emploi de nouveaux programmes nommés interfaces utilisateur. Aujourd'hui des logiciels d'analyse de textes et de manipulation de données hiérarchisées facilitent la recherche sur le réseau Internet ou dans de grandes bases de données, car la plupart des sites informatiques sont organisés en «pages», une structure simple et familière, mais inutilement restrictive. Dans l'avenir, cette présentation sera remplacée par des structures qui montreront les informations de plusieurs points de vue simultanés. Imaginons qu'Alice, dans la Creuse, se connecte au réseau afin de savoir quels types de bulbes comestibles, tels l'ail ou l'oignon, elle peut planter dans son jardin. Les réponses à sa question se trouvent certainement sur le réseau, mais où?

Alice dispose aujourd'hui de plusieurs techniques pour s'y retrouver, mais aucune n'est parfaitement satisfaisante. Elle peut d'abord demander à des amis s'ils connaissent des sites inté-

ressants. Elle peut aussi feuilleter des index du réseau, qui sont aujourd'hui de deux types : soit des listes construites manuellement, qui classent les sites par type de contenu, soit des moteurs de recherche qui repèrent rapidement des mots clés dans les index du réseau.

Grâce à leurs employés qui répertorient quotidiennement des centaines de documents, des sites, tels *Yahoo* et *Nomade*, proposent aux utilisateurs une table des matières. Pour y rechercher l'information, on choisit dans un menu (voir la figure 1) la catégorie qui semble la plus proche de celle que l'on cherche et l'on reçoit un sous-menu plus spécialisé ou une liste de sites dont le personnel de la société pense qu'ils appartiennent à cette catégorie. Tou-

tefois l'emploi de ces tables des matières n'est pas très simple, car les catégories ne sont pas toujours mutuellement exclusives : Alice doit-elle choisir la rubrique «Loisirs», ou bien la rubrique «Écologie», par exemple? Quel que soit son choix, le menu disparaît de l'écran, de sorte qu'en cas d'échec elle doit refaire toute la recherche, et répondre à toutes les questions posées par l'interface de recherche. Si l'un de ses choix n'aboutit pas (si la rubrique «Écologie» ne contient pas l'information cherchée, par exemple), elle doit repartir à zéro et en essayer une autre. Si l'information désirée est enfouie dans les profondeurs de la hiérarchie ou n'existe pas, la procédure peut être fastidieuse, voire exaspérante.



1. LE SYSTÈME de mise en page utilisé sur la plupart des sites du réseau impose des contraintes superflues aux services, tels *Yahoo* ou *Nomade*, qui tentent de structurer le contenu du réseau, afin de faciliter la recherche d'informations. Un utilisateur qui désire connaître les type de bulbes comestibles (ail, oignon, etc.) qu'il peut planter en automne, par exemple, peut choisir une catégorie. Plus il précise sa demande, plus il lui est difficile de se rappeler quelles catégories potentiellement utiles il n'a pas encore explorées.



Marti A. Hearst, Xerox Parc

2. DES PROGRAMMES de recherche facilitent l'exploration du réseau en créant des livres virtuels qui contiennent les pages trouvées, ainsi que des liens vers des sites. Ces livres qui stockent les résultats des recherches sont placés sur des étagères virtuelles, où l'on peut les reprendre ultérieurement. Ils peuvent également guider l'utilisateur pour la recherche de nouvelles informations.

La lenteur de la pensée

Durant la dernière décennie, les recherches sur la visualisation de l'information ont engendré plusieurs techniques efficaces pour l'affichage d'ensembles de données abstraites sous une forme plus propice à l'exploration. L'une des stratégies consiste à transformer le travail demandé à l'utilisateur, en remplaçant les processus lents et intellectuellement exigeants – telle la lecture – par des processus perceptifs plus rapides, comme la reconnaissance de formes. Ainsi on compare plus facilement les barres d'un histogramme que les nombres d'une liste. La couleur, également, facilite la détection des mots ou objets dans un océan de mots ou d'objets.

Une autre stratégie consiste à exploiter l'illusion de profondeur que peuvent donner les écrans : quand l'affichage donne une image bidimensionnelle d'un monde à trois dimensions, la perspective, le recouvrement et les ombrages fournissent des indices perceptifs qui clarifient les relations entre des documents qui formeraient une mosaïque indéchiffrable dans une page à deux dimensions ; on peut mettre au premier plan les informations les

plus intéressantes et rejeter à l'arrière-plan ou à la périphérie de l'écran les informations secondaires ou celles qui ont conduit aux informations cherchées. Ainsi, l'utilisateur garde une vision structurée de sa recherche.

Un tel environnement facilite l'exploration. Les utilisateurs peuvent dénicher des résultats partiels qu'ils utiliseront ultérieurement, mieux formuler leurs demandes, reprendre des voies qu'ils avaient jugées inintéressantes ou envisager leurs demandes sous un angle nouveau. Aujourd'hui Alice peut avoir cette démarche si elle prend des notes à mesure qu'elle utilise *Yahoo*, mais un prototype conçu par mes collègues du Centre de recherches Xerox vise à accroître l'efficacité de ces activités intelligentes.

Nommée *Information Visualizer*, ce programme dessine un arbre tridimensionnel (voir la figure 3) qui relie chaque catégorie à toutes ses sous-catégories. Si Alice cherche l'arbre *Yahoo* pour «jardins», les six zones de *Yahoo* dont «jardin» ou «jardinage» sont une sous-catégorie s'affichent simultanément. En faisant tourner l'arbre, Alice peut amener chacune de ces catégories au premier plan pour découvrir où elle conduit. Lorsqu'un chemin aboutit à

une impasse, il suffit de cliquer une fois pour retrouver les voies non explorées.

Lorsque Alice trouve une information intéressante, elle peut utiliser le programme pour la stocker dans un livre virtuel, avec les mots clés qui ont été utilisés ; puis elle peut placer ce livre sur une étagère virtuelle, où il est visible et nettement étiqueté. Le week-end suivant, par exemple, elle peut reprendre sa recherche à l'endroit où elle l'a laissée : il lui suffit d'ouvrir le livre et de relancer la recherche à partir d'une de ses pages.

Notre interface ne pourrait servir à structurer tout le réseau : les sites nouveaux apparaissant bien trop rapidement pour être indexés à la main, la proportion de sites répertoriés par *Yahoo* (ou par tout autre service) diminue de jour en jour. En outre, certains sites, tel celui des quotidiens, contiennent des articles sur tant de sujets qu'ils ne sont souvent rangés que dans quelques-unes des catégories possibles.

Les moteurs de recherche tels que *Excite* et *AltaVista* sont bien plus généralistes, mais c'est là leur faiblesse. Si Alice fournissait à *Excite* la chaîne de mots clés «ail, oignon, automne, jardin croissance», elle récupérerait 583 430 pages, ce qui (au rythme de deux minutes par page) lui demanderait plus de deux années ininterrompues de feuilletage. Plus généralement, toute recherche exhaustive procure inévitablement des listes encombrées d'informations superflues, tandis que les recherches plus ciblées risquent de passer à côté de nombreuses pages intéressantes.

Structurer les résultats de recherche

La concision des formulaires d'entrée de la plupart des services de recherche aggrave le problème, parce qu'elle incite les utilisateurs à faire des demandes lapidaires et nécessairement vagues. On devrait encourager l'emploi d'opérateurs logiques tels que AND, OR et NOT pour spécifier les mots qui doivent (ou ne doivent pas) figurer sur les pages affichées. Cependant cette notation booléenne intimide ou dérout de nombreux utilisateurs, et même les spécialistes n'obtiennent que des réponses dont la qualité dépend des termes utilisés dans leur demande. Une meilleure solution consiste à