

Nathalie Nadia Mitev*

Department of Information Science, The City University, Londres

L'AUTOMATISATION DES CATALOGUES

INTERACTION UTILISATEUR/SYSTÈME

L'AUTOMATISATION des bibliothèques a débuté en France (1, 2, 3) depuis une quinzaine d'années, suivant le mouvement amorcé dans le monde anglo-saxon. De façon assez uniforme, les premières fonctions de la bibliothèque à être informatisées ont été, et sont encore, le catalogage et le prêt. Ce choix, pour des raisons évidentes, a été un choix de gestion, d'une part afin d'éviter la répétition des tâches et la redondance des fichiers, d'autre part afin d'améliorer l'efficacité du service. Outre-Manche et Outre-Atlantique, ces deux fonctions, considérées comme les fonctions bibliothéconomiques primordiales, sont les premières qui sont venues à l'esprit des bibliothécaires lorsqu'il s'est agi d'automatiser leur travail. D'autres applications étudiées en second lieu concernent la gestion des acquisitions et des commandes, le bulletinage des périodiques, etc. En règle générale, l'interrogation/consultation interactive de catalogues en ligne par le public, bien qu'étant une application importante, n'est encore envisagée qu'en dernier lieu ou après coup.

En effet, le module de consultation est rarement conçu en tant que tel, mais comme faisant partie d'autres modules servant d'abord le (la) bibliothécaire pendant qu'il (elle) catalogue, ou gère le prêt.

La plupart du temps, une fois testés les logiciels de catalogage ou de circulation, un terminal — et souvent au départ un seul — est mis à la disposition des utilisateurs un peu par hasard. Il est alors installé face au personnel, et tourne donc le dos au public; le bibliothécaire intervient comme médiateur/intermédiaire si une demande est formulée.

* L'auteur, française, a étudié et travaillé en Grande-Bretagne pendant ces quatre dernières années sur plusieurs de projets de recherche du British library research & development department; l'un d'eux concernait la conception, la réalisation et l'évaluation d'un catalogue interactif prototype.

Les bibliothèques américaines, s'automatisant vers la fin des années 70, n'avaient absolument pas prévu que leurs catalogues en accès direct allaient être utilisés de façon aussi intensive par les lecteurs, et les bibliothécaires ont été dans l'ensemble très surpris par ce phénomène.

CONTEXTE GÉNÉRAL

Avant d'entrer dans les détails des catalogues interactifs, il est peut-être nécessaire de clarifier le contexte global dans lequel ils se situent. Ils se distinguent de plusieurs manières: le type de fichier, le type d'organisme dans lequel ils sont utilisés, les utilisateurs concernés et le mode d'utilisation.

Fichiers documentaires et organismes

On peut dire qu'il existe couramment au moins trois sortes de fichiers documentaires automatisés, utilisés dans des contextes différents par divers types d'utilisateurs:

- les bases de données bibliographiques et les banques de données (qui ont souvent l'étiquette « information scientifique et technique ») utilisées dans le secteur industriel et commercial et dans le secteur universitaire;
- la télématique grand public, par exemple l'expérience CLAIRES à la Mairie de Grenoble (4), le vidéotex ou le Minitel (5);
- les catalogues interactifs, qui permettent de retrouver les ouvrages physiquement disponibles dans une bibliothèque donnée.

Par *fichier documentaire*, on entend fichier « pour se documenter », servant à une recherche documentaire. Cela comprend et dépasse la notion de fichier bibliographique au sens strict du terme. Néanmoins, le terme de recherche documentaire doit être défini de manière précise: de nombreuses situations de travail comportent un aspect recherche d'information sur ordinateur pour effectuer toutes sortes de tâches. Par exemple, un chimiste ou un homme d'affaires utilisent certains types de fichiers automatisés pour rechercher des données afin de les traiter, de les modifier, d'en créer de nouvelles, de les manipuler de façon active. Mais leur

objectif principal n'est pas une recherche d'information en tant que telle. Une recherche documentaire est une consultation « passive » de données bien établies, sans modification de ces données, dans le but de répondre à une *quête de connaissance* (6). Des fichiers documentaires existent pour aider un groupe d'utilisateurs à satisfaire ses besoins documentaires.

Les modes d'utilisation

Ces trois types de fichiers documentaires diffèrent par les catégories d'utilisateurs qu'ils servent, et par la façon dont ils sont utilisés.

Les bases de données scientifiques et techniques sont conçues pour répondre aux besoins documentaires des ingénieurs, techniciens, cadres, universitaires, chercheurs, etc. En général elles sont interrogées par un professionnel de l'information (documentaliste, ingénieur spécialisé) qui sert d'intermédiaire/médiateur entre ces fichiers et le groupe d'utilisateurs; ceci pour des raisons d'efficacité et de rapidité: un ingénieur n'a pas le temps d'apprendre à se repérer dans les nombreuses bases existantes ni à manipuler les divers logiciels d'interrogation et les caractéristiques des langages booléens de commande, leurs codes et syntaxe, etc., à moins d'être un utilisateur suffisamment régulier et motivé.

Les fichiers télématiques, annuaire électronique et autres, sont en principe conçus pour le « grand public » qui est en contact direct avec le terminal, le logiciel de dialogue et les bases/banques de données disponibles (qui foisonnent sur le Minitel). Les catalogues interactifs sont comparables aux systèmes télématiques dans le sens où ils sont utilisés, ou devraient pouvoir être utilisés, *directement* par les utilisateurs finals, donc sans intermédiaire. Une des différences, et non négligeable, est que ces fichiers renvoient à des ouvrages accessibles *immédiatement*. Des informations sur le statut du livre sont affichées à l'écran (s'il est déjà emprunté, si on peut le réserver en ligne, etc.), ce qui crée une relation particulière entre l'utilisateur et le système.

L'interaction directe entre le catalogue en ligne et l'utilisateur peut inquiéter certains: « (*les biblio-*

*thécaires) souhaitent-ils conserver le type de médiation directe qu'ils connaissent, la relation à l'usager — ou se sentent-ils prêts à la transposer sur d'autres supports ? » (4). Brigitte Guyot ajoute elle-même que les deux approches ne sont pas incompatibles. On peut trouver des utilisateurs qui préfèrent interroger une base de données plutôt que de passer par le bibliothécaire (7); on peut alors dire qu'intervient la notion de *self-service* (qui n'est pas une nouveauté, car tout catalogue est utilisé directement par les utilisateurs, sans intermédiaire).*

La facilité d'usage

En tout état de cause et de façon générale, les utilisateurs, leurs tâches, besoins et comportements sont rarement pris en compte lors du processus de conception et de réalisation de systèmes informatiques, sinon simplement en fin de parcours. Un processus itératif d'expérimentation et d'évaluation est fortement recommandé (8), non seulement une fois le produit achevé, mais avant, pendant et après sa conception. Un produit informatique n'est d'ailleurs jamais totalement achevé... Dans le cas des catalogues interactifs, il est intéressant, avant même toute introduction d'ordinateurs, d'examiner ce que les utilisateurs font quand ils utilisent un catalogue de bibliothèque, comment ils repèrent des livres sur les rayons (9).

Il est important de connaître les besoins, les préférences, les réactions des utilisateurs, mais ce qu'ils expriment n'est pas suffisant. Et, plus que les utilisateurs eux-mêmes, c'est *l'utilisation* d'un système par la population visée qui devrait être une priorité d'étude et d'évaluation régulière.

Peu d'études ont été faites sur les comportements de recherche d'information des utilisateurs, surtout non-experts, lorsqu'ils sont confrontés à des outils documentaires. Ces comportements, qui sont aussi fortement liés à ces outils, permettent de voir dans quelle mesure les outils sont *utilisables*. Leur *facilité d'usage* (10) devrait être examinée sérieusement et il est important de prendre en compte ce critère lorsque l'on s'intéresse au taux d'acceptabilité et de succès d'un système d'information.

LES CATALOGUES INTERACTIFS

Du point de vue de l'utilisateur, les catalogues ont pris différentes formes qui se sont succédé de façon plus ou moins linéaire, du catalogue sur fiches cartonnées classées dans des tiroirs, au catalogue sur microfiches accessible par l'intermédiaire de lecteurs de microfiches, avant d'être accessible par l'intermédiaire d'un terminal de consultation. La plupart des bibliothèques américaines ne sont pas passées par le stade de la microfiche, à l'inverse des bibliothèques britanniques.

Autant aux Etats-Unis qu'en Grande-Bretagne existent depuis un certain nombre d'années des fichiers *sur machine* correspondant aux fonds documentaires des bibliothèques nationales ou des coopératives régionales de bibliothèques partageant le catalogage de leur stock et créant ainsi des catalogues collectifs informatisés (LOCAS, BLCMP, SWALCAP en Grande-Bretagne, RLIN, WLN, OCLC aux Etats-Unis). Les bibliothèques utilisant ces catalogues partagés évitent ainsi de répéter les tâches de catalogage. Ces catalogues sur machine peuvent être déchargés localement ou achetés. le format d'échange MARC, correspondant aux règles de catalogage anglo-américaines AACR2, donne la possibilité d'échanger facilement les bandes magnétiques.

L'automatisation des bibliothèques est, en général, développée par trois types d'organismes:

- par l'organisme utilisateur, c'est-à-dire une bibliothèque de grande importance ou un réseau de bibliothèques, par exemple l'University of California, Division of library automation qui a développé un système appelé MEL-VYL (11);

- par un consortium ou une coopérative nationale ou régionale de catalogage partagé (voir ci-dessus);

- par une compagnie commerciale, par exemple GEAC, IBM, MATRA, etc. Ces compagnies souvent ne traitaient au départ que la gestion du prêt; elles fournissent maintenant de plus en plus de systèmes « intégrés », c'est-à-dire permettant la plupart des fonctions bibliothéconomiques mentionnées ci-dessus.

Du papier à l'OPAC

Mise à part sa fonction d'aide au catalogage, l'ordinateur permet aussi d'éditer les catalogues soit sur fiches papier (Library of Congress aux Etats-Unis) soit sur microfiches (LOCAS à la British library) pour les bibliothèques individuelles appartenant à ces réseaux.

Progressivement, ces fichiers bibliographiques sur machine ont été considérés comme des bases de données — donc accessibles directement — et non plus seulement comme des outils de création de catalogues traditionnels. Peu de bibliothèques avaient prévu sinon imaginé la « fermeture » des catalogues traditionnels; ainsi les bibliothèques universitaires de l'Ohio State University ont finalement « fermé » leur catalogue sur fiches papier en 1982 (12), ce qui n'avait absolument pas été prévu quand l'automatisation y a débuté à la fin des années 70. Une production de catalogues sur support traditionnel (fiche ou microfiche) existe encore souvent parallèlement à l'accès direct.

Aux Etats-Unis, les systèmes de catalogues interactifs sont déjà

nombreux: il existe, ainsi qu'au Canada, une cinquantaine de systèmes différents utilisés dans plusieurs centaines de bibliothèques (12, 13, 14). Ils commencent à se développer en Australie et en Europe, surtout en Grande-Bretagne (15), mais aussi en Hollande, en Belgique, en Suisse, en Allemagne et en France. leur nombre devrait atteindre bientôt la centaine dans le monde entier. Ce développement est relativement récent, et la plupart de ces systèmes datent du début des années 80. Pour la première fois en 1985, l'*Annual review of information science and technology* qui fait le point sur les principales questions en science et technologie de l'information, a publié un article sur les catalogues interactifs (12). Ils sont souvent dénommés par un acronyme maintenant accepté et couramment usité: OPAC pour *Online public access catalogue*.

Ces catalogues interactifs ou OPACs sont facilement utilisés, sans intermédiaire humain, sans documentation ni manuel d'utilisation et sans formation, par une population type grand public de plus en plus hétérogène et exi-

geante dans des bibliothèques publiques, universitaires, régionales et nationales. Plusieurs enquêtes aux Etats-Unis (16) montrent qu'ils sont très appréciés par leurs utilisateurs.

OPACs : des outils de recherche documentaire

Les principes de conception des catalogues interactifs existants peuvent être examinés sous deux angles: les techniques de recherche d'information et les modes de dialogue (cf. références 7, 17 et 18).

Ces deux aspects ne sont pas indépendants: les techniques de recherche d'information utilisées ont une influence certaine sur la conception d'un mode de dialogue. Réciproquement, on ne peut pas parler de communication homme/machine d'un système documentaire sans tenir compte de l'efficacité des techniques de recherche d'information; après tout, le public a besoin de systèmes faciles à utiliser, qui *marquent*, c'est-à-dire de systèmes dont les techniques de recherche d'information soient efficaces, et

ce d'autant plus que les utilisateurs ne sont pas des experts. On peut classer les catalogues à accès direct en trois « générations » (19).

Première génération

Elle est issue soit des systèmes de prêt (comme les premières versions de GEAC), soit des systèmes de catalogue déjà mentionnés. Ces derniers contiennent en général une description bibliographique assez complète (format MARC ou dérivé). Quant aux premiers, ils ne stockent qu'une description minimale des ouvrages (titre tronqué, pas de mots-clés ni de termes d'indexation), et on entre en général le numéro d'inventaire du livre ou le numéro de carte de lecteur, par lecture optique. On peut retrouver un ouvrage donné en entrant un « acronyme » ou une « clé auteur/titre » extrêmement abrégée comme par exemple « AT/HUGO, MISE » ou « STENDHROUG, NOI ».

Accès par phrases

L'accès à l'information contenue dans ces catalogues automatisés de première génération simule la recherche d'information dans les catalogues traditionnels (tiroirs à fiches ou microfiches); celle-ci est influencée par le concept bibliothéconomique classique suivant : un catalogue est la représentation *unique* des ouvrages; chaque ouvrage doit pouvoir être identifié de façon unique et non équivoque. Ce concept va de pair avec l'idée qu'un catalogue sert d'abord à retrouver des ouvrages déjà connus. D'où l'usage de *phrases*: c'est-à-dire que, si l'on prend l'exemple du titre, l'on doit se souvenir exactement de son libellé. *L'ordre* des mots du titre est très important. Une fois le titre entré, le système affiche sur l'écran une liste de titres par ordre alphabétique autour du titre entré. Si le titre désiré n'est pas dans cette liste, l'utilisateur est laissé à lui-même, et la seule chose qu'il peut faire est de se « promener » dans cette liste de haut en bas, en suivant l'ordre alphabétique. Les systèmes de sélection par écran tactile (comme les écrans du système ALS) sont aussi basés sur cette méthode d'interrogation des fichiers par sélection alphabétique de phrases. C'est un accès *pré-coordonné*.

Ce type d'accès est relativement efficace pour retrouver des données exactes; mais l'utilisateur doit

déjà connaître précisément un ou plusieurs éléments des ouvrages qu'il recherche. Par contre, quand il s'agit de faire une recherche par sujet, l'approche pré-coordonnée est limitée (cf. encadré : système LCS), car elle ne fournit pas suffisamment de points d'accès.

En termes de dialogue, ces systèmes offrent en général une approche par *menus* (cf. l'exemple de DOBIS) : l'utilisateur choisit un critère de sélection parmi une liste de champs (auteur personne physique, collectivité auteur, titre, sous-titre, indice de classification, etc.) correspondant aux index informatiques par phrases, construits sur chaque champ de la notice bibliographique.

Seconde génération

La deuxième génération est influencée par les systèmes d'interrogation en ligne des bases de données bibliographiques traditionnelles comme Pascal sur Questel ou Inspec sur Dialog. Ces

bases de données contiennent des références à des articles de périodiques et très peu de monographies. La plupart comportent des résumés et un grand nombre de termes d'indexation (souvent il y a un thésaurus ou vocabulaire contrôlé) représentant beaucoup de texte.

Accès par mots

Afin de retrouver des mots de ce texte, on doit construire des index à ces bases de données contenant des *mots* individuels. L'accès y est *post-coordonné*.

En ce qui concerne la recherche par sujet, ce principe appliqué dans les catalogues interactifs permet de retrouver n'importe quel mot de la description bibliographique d'un ouvrage, et l'utilisateur n'a pas besoin de se souvenir de phrases exactes. Le nombre de points d'accès est énormément accru, ce qui favorise les recherches par sujet. Rien d'étonnant à cela : les bases de

Recherche par sujet sur le système LCS* (Ohio State University)

/SIS/communication en physique

10	1	communication en administration
11	3	communication en géographie
12	1	communication en hydrologie
13	1	communication en médecine
14	1	communication en organisations
15	1	communication en politique
16	3	communication en sciences
17	2	communication en technique
18	1	communication en urbanisme

Input : communication en physique

page 2 de 3 pour page suivante entrer ps ou numéro de page

* le système donne la liste alphabétique des vedettes matières autour de l'expression entrée. Traduit et adapté de (13).

Menu de DOBIS/Leuven, IBM (13)

Searching

System files

- 1 Names
- 2 Titles
- 3 Subjects
- 4 Publishers
- 5 Classification
- 6 ISBN/ISSN
- 7 LC card numbers
- 8 Other numbers
- 9 Document numbers

Enter number or code

1_

Local files

- 10 Names
- 11 Titles
- 12 Subjects
- 13 Shelf list numbers
- 14 Copies, volumes
- 15 Document numbers

e end

R

données scientifiques et techniques étant conçues pour répondre surtout à des besoins de recherches par sujet.

Le mode de dialogue dans les systèmes à accès post-coordonné est le langage booléen de commande (cf. encadré: système TOMUS). Ces langages d'interrogation permettent de combiner les mots des index en utilisant des opérateurs booléens.

Avantages et inconvénients

Il est devenu de plus en plus évident (10, 19) que ces deux premières générations ne sont pas entièrement satisfaisantes. Elles ont chacune des avantages et des inconvénients.

Du point de vue des techniques de recherche d'information, la pré-coordination est utile lorsqu'il s'agit de retrouver des données exactes, mais échoue trop facilement dès que manque un élément. Elle est plus appropriée pour retrouver des ouvrages déjà connus. Il est plus facile de rechercher un auteur ou un titre en utilisant l'accès par phrases plutôt qu'en utilisant l'approche booléenne: combiner les deux mots « Guerre » et « Paix » avec un opérateur booléen *ET* nous donnera énormément d'éléments non pertinents (« Les efforts de paix dans la guerre du Proche-Orient »). Par contre, l'approche booléenne est plus pertinente pour les recherches par sujet.

Les catalogues doivent pouvoir répondre à ces deux types de recherche d'information. Il s'est avéré que les utilisateurs de catalogues interactifs font des recherches par sujet au moins autant, sinon plus, que des localisations d'ouvrages déjà connus (20), contrairement à ce que croyaient beaucoup de bibliothécaires anglais et américains.

Les méthodes utilisées (pré-coordonnée et post-coordonnée) ont eu des conséquences sur le mode de dialogue utilisé (menu, langage de commande) et cela se retrouve au niveau de la communication: le langage utilisé sur l'écran reflète les attitudes des concepteurs des systèmes. Les premiers OPACs emploient souvent du vocabulaire bibliothéconomique spécialisé que tout un chacun ne connaît pas, comme « vedette matière », « collectivité auteur », correspondant aux noms de champs des notices. Dans les systèmes de deuxième génération, on trouve souvent du jargon booléen incompréhensible pour

le non-initié, comme « *Utilisez '* comme symbole de troncature* ». On trouve souvent aussi du jargon informatique (« *Control Z pour menu* », « */Enter/ + 'Mot'Esc/* »), ce qui n'améliore pas la communication homme/machine.

La méthode par menus passe pour plus appropriée aux utilisateurs inexpérimentés ou occasionnels, tandis que celle des langages de commande conviendrait aux utilisateurs expérimentés ou réguliers.

Un premier problème se pose: comment passe-t-on d'un stade à l'autre? A l'Université de Californie (11), MELVYL propose les deux modes de dialogue à ses utilisateurs: mais il s'est avéré qu'une énorme majorité d'entre eux ne s'aventure jamais dans le langage de commande.

Un autre problème tient à ce que l'interface par menus est très contraignante et ne donne pas à l'utilisateur la possibilité de se servir d'outils de recherche beaucoup plus puissants. D'un autre côté, on ne peut pas non plus forcer tous les utilisateurs à apprendre les langages booléens de commande: non seulement ils ne sont pas évidents à comprendre, mais être obligé d'exprimer une recherche d'information sous

forme d'équation booléenne n'est pas un acte naturel (sinon les intermédiaires spécialistes de l'interrogation de bases de données n'existeraient pas...). Même le concept de recherche sur des phrases et des mots, qui peut paraître évident pour des spécialistes, n'est pas simple à expliquer sur un écran.

Vers une troisième génération

Pour rendre ces systèmes accessibles et utilisables par le plus grand nombre possible de gens, il faut améliorer à la fois les techniques de recherche d'information et les interfaces utilisateurs.

De plus en plus de systèmes de seconde génération offrent maintenant les deux formes d'accès, par phrases et par mots, mais de façon assez « crue »: l'utilisateur est obligé dès le début de choisir lui-même entre ces deux formes d'accès, et il est clair que la plupart des utilisateurs ne sont pas en mesure de comprendre leurs avantages et inconvénients. Quelques systèmes, que l'on peut appeler de troisième génération, emploient différentes techniques.

Les techniques de recherche

PaperChase (21) et surtout CITE (22), développé à la National library of medicine aux Etats-

TOMUS Catalogue de la New York public library (CATNYP)*

Pour chercher un sujet, entrer **TROUVER SUJET** (ou juste **TS**), puis le sujet; par exemple, **TROUVER SUJET RÉFUGIÉS**, ou **TS DANSE**.

Pour chercher un auteur, entrez **TROUVER AUTEUR** (ou juste **TA**), puis l'auteur; par exemple, **TROUVER AUTEUR MARCEL PROUST** ou **TA ALBERT CAMUS**.

Pour chercher un titre, entrez **TROUVER TITRE** (ou juste **TT**), puis le titre; par exemple, **TROUVER TITRE CONTES DU CHAT PERCHE** ou **TT HISTOIRE DE LA DANSE**.

Pour trouver un ouvrage avec l'auteur et le titre, entrez **TROUVER AUTEUR**, l'auteur, **ET TITRE**, puis le titre; par exemple, **TROUVER AUTEUR HUGO ET TITRE MISÉRABLES**.

Après que vous ayez entré une de ces commandes, ou des instructions à **CATNYP**, pressez la touche **RETURN**. **CATNYP** cherchera le catalogue et vous dira combien d'éléments il a trouvés. Pour voir ces documents, entrez **VISUALISER** (ou juste **V**) et pressez la touche **RETURN**.

Si vous faites une erreur en tapant, pressez la touche marquée **BACK** pour reculer, puis retaper.

Pressez toujours la touche **RETURN** quand vous avez fini de taper.

Pour plus d'information, entrez **AIDE** (ou juste **A**), puis pressez la touche **RETURN**.

* Traduit et adapté de (13).

R

Unis, utilisent une technique de recherche d'information mise au point par des chercheurs en informatique documentaire mais qui, jusque là, n'avait pas été intégrée à des systèmes opérationnels. Dans CITE, l'utilisateur entre sa recherche par sujet en « langage naturel » : par exemple : « les maladies du sang chez les enfants et nourrissons ». CITE effectue une recherche combinatoire booléenne sur les mots de cette expression, sur les termes morphologiquement apparentés et sur les termes du thésaurus médical (*Medical subject headings*). Les références des documents trouvés sont montrées à l'utilisateur et le système lui demande d'indiquer celles qui sont pertinentes et les termes jugés importants. Le système modifie et recommence sa recherche, présente les nouveaux résultats à l'utilisateur, reformule en fonction de ses jugements de pertinence, etc., jusqu'à ce que l'utilisateur soit satisfait des résultats et décide d'arrêter. Cette méthode est appelée le « *feedback* de pertinence ». CITE a l'avantage de pouvoir combiner cette méthode avec l'usage d'un thésaurus important et bien défini (qui correspond à l'*Index medicus*), ce qui n'est pas le cas de beaucoup de catalogues. De plus, le domaine couvert est bien défini et limité et les utilisateurs de CITE sont des médecins ou spécialistes médicaux fortement motivés qui connaissent bien la terminologie médicale.

Okapi (7) effectue des opérations booléennes *automatiquement* afin que l'utilisateur n'ait pas à formuler d'équations compliquées. Il a aussi perfectionné la méthode de combinaison booléenne en y greffant un algorithme de pondération automatique : les termes d'une

expression entrée en « langage naturel » sont combinés de différentes façons selon le calcul des « poids » attribués à chaque terme en fonction inverse de leur fréquence. Cet algorithme est appelé automatiquement quand un *ET* sur plus de deux mots ne produit pas de résultat. Les références trouvées sont présentées à l'utilisateur dans l'ordre décroissant de « ressemblance » avec l'expression entrée. L'exemple de l'encadré donne les titres comprenant les mots suivants, dans cet ordre : « *motivation, work, industry* », « *work, psychology, industry* », « *psychology, industry* », « *motivation* » étant le plus rare, donc ayant le poids le plus élevé. Les documents avec le plus de mots et de poids élevé apparaissent les premiers.

Okapi utilise aussi les deux méthodes d'accès par mots et par phrases, alternativement selon le résultat en suivant des « arbres » décisionnels relativement simples. Par exemple, si l'utilisateur entre un titre avec plusieurs mots, le système essaye d'abord une recherche sur phrase; si cette procédure échoue, il combine les mots individuellement avec un *ET*; en cas d'échec et s'il y a plus de deux mots, il tentera la recherche booléenne pondérée qu'on vient de décrire.

Les interfaces utilisateurs

Un équilibre difficile doit être trouvé entre les deux extrêmes suivants :

— exposer tous les détails techniques de l'informatique documentaire, ce qui rend les systèmes utilisables uniquement par des professionnels ou quasi-professionnels;

— rendre un système « invisible ». On trouve dans des brochures publicitaires le qualificatif « transparent » attribué à des logiciels censés être faciles à utiliser. Le terme « invisible » semble plus juste, car il exprime mieux ce que certains systèmes font : tout se passe « derrière » l'écran, et on n'explique pas à l'utilisateur ce qui se passe. L'invisibilité totale est inefficace car, ne permettant pas à l'utilisateur de comprendre ce qui se passe, elle ne lui permet donc pas de modifier sa stratégie en cas d'échec.

La communication homme/machine dans les systèmes documentaires pour le grand public englobe plus que la seule présentation de l'information sur l'écran. Il ne s'agit pas seulement « d'embellir » ou de rendre convivial un système de recherche d'information pré-existant qui marche plus ou moins; l'interaction avec l'utilisateur doit être prise en compte dès la conception des méthodes d'accès à l'information; celles-ci doivent être envisagées avec l'objectif de faciliter la tâche de recherche d'information de l'utilisateur, en rendant le système le plus « visible » possible.

Okapi et ses fans

Okapi a tenté d'améliorer la clarté, la lisibilité et la simplicité des messages, des instructions et de la présentation de l'information. Par exemple les principales touches de fonction sur le clavier sont en couleur; un langage courant est préféré à du jargon (« *Pressez la touche verte* », « *Livre* » au lieu de « *Enregistrement* »).

Mais Okapi tente aussi d'aider le plus possible ses utilisateurs en cas d'échec. Le problème de l'échec est particulièrement important pour la communication homme/machine : les utilisateurs des OPACs ne sont généralement pas très persévérants et se découragent vite. Okapi essaye de ne pas les laisser (les lasser !) avec des messages du type « *Résultat négatif pour votre recherche* », sans essayer de lui-même, si possible, une autre méthode ou sans la suggérer. C'est bien le principe sous-jacent à ses arbres décisionnels et à sa recherche combinatoire pondérée.

Les arbres décisionnels permettent aussi à Okapi d'offrir à l'utilisateur le choix entre seulement deux critères de sélection, une

Recherche par sujet dans Okapi Recherche booléenne combinatoire pondérée (7)

The computer will look for book(s) described by as many as possible of the word(s) you type. Please enter word(s) or a short phrase which describe your subject :

the psychology of motivation at work in industry

Searching for book(s) described by ALL these words

NO BOOK FOUND

« psychology » : 1454 « motivation » : 72 « work » : 1221

« industry » : 1587

Searching for SIMILAR books

GREEN KEY to look at the books found

(The most similar books will appear first) —

recherche pour un « *Book about something* » ou une recherche pour un ouvrage spécifique déjà connu, au lieu de lui présenter une liste exhaustive de tous les champs interrogeables comme dans le menu de DOBIS (cf. p. 241). Les arbres permettent d'interroger plusieurs champs à la fois, ou d'interroger des champs dans un certain ordre, selon ce que l'utilisateur a entré et selon les résultats. Par exemple, si quelqu'un entre trois mots ou plus comme nom d'auteur, le système va considérer qu'il peut s'agir d'une collectivité auteur.

Les systèmes documentaires doivent aussi guider une stratégie de recherche d'information employée fréquemment par les utilisateurs : le « butinage » (« *browsing* »). L'accès par phrases des OPACs de première génération fournit la possibilité de « butiner » dans les index par ordre alphabétique, ce qui est un énorme avantage. Mais imposer l'ordre alphabétique n'est peut-être pas le meilleur moyen de permettre ce butinage. Il est beaucoup plus difficile d'aider ce genre de stratégie dans les systèmes booléens. La recherche combinatoire pondérée utilisée par Okapi a l'avantage de montrer à l'utilisateur un nombre en général assez grand

de documents, suivant un principe de « ressemblance » qui est peut-être plus utile que l'alphabet : on regarde les documents un par un, en commençant par ceux qui ont le plus de mots en commun avec le sujet exprimé, mais les documents suivants peuvent aussi avoir de l'intérêt et permettent en général d'« élargir » le sujet initial.

Pleins feux sur l'usager

Les catalogues interactifs peuvent jouer un rôle essentiel d'aiguillage et de familiarisation d'une population importante à l'informatique documentaire. Dans l'idéal, références bibliographiques, catalogues et banques factuelles grand public devraient être accessibles sur un seul terminal. La bibliothèque publique de Pikes Peak au Colorado (23) s'approche de cet idéal en ce qui concerne l'accès aux ouvrages de la bibliothèque et à l'information communautaire. Par ailleurs, une fois qu'un catalogue est chargé sur une machine, il ne coûte rien, en termes de frais de connection, télécommunications, serveurs, etc., de le rendre accessible au public de cette bibliothèque, mis à part le coût initial des terminaux de consultation.

Il est important d'être attentif à la facilité d'usage des fichiers automatisés, surtout si l'on veut qu'ils soient utilisables par un grand nombre de gens. Il semble que l'utilisateur moyen se situe au milieu d'une myriade de fichiers et systèmes d'information de toutes sortes, sans savoir où se diriger, par où commencer. Le schéma devrait être inversé : l'information se situant au milieu et toutes sortes d'utilisateurs autour, y accédant.

Cependant, la plupart des catalogues interactifs existants sont encore relativement primitifs en ce qui concerne leurs interfaces utilisateurs; un travail d'élaboration de leurs techniques de recherche d'information doit être effectué, avec comme but principal l'amélioration de la qualité de leur communication homme/machine. Il y a plusieurs équipes de recherche en Grande-Bretagne et aux Etats-Unis qui travaillent sur différents aspects de conception des OPACs; ainsi une équipe anglaise travaille sur des algorithmes d'approximation de l'orthographe des noms (pour résoudre les problèmes fréquents de fautes d'orthographe dans les noms), sur des algorithmes visant à rassembler les mots de même racine, et sur la manière de générer un

R

nombre limité de synonymes. Karen Markey et son équipe à l'OCLC en Ohio explorent l'usage de la CDD (Classification décimale Dewey) pour aider les recherches par sujet. Une autre équipe essaye d'enrichir la description par sujet des notices bibliographiques en utilisant PRECIS (*PREserved context index system*), d'autres en intégrant les sommaires et index des monographies au catalogue.

Néanmoins, on ne peut ajouter des moyens sophistiqués d'interrogation sans *évaluer* leur pertinence et leur facilité d'usage.

Un autre domaine de recherche de première importance s'ouvre également, l'étude de l'utilisation des OPACs et leur évaluation.

Une des premières méthodes utilisées a été l'analyse des transactions automatiques des interrogations. Les OPACs étant des systè-

mes locaux, il est facile d'avoir un programme qui enregistre automatiquement les sessions de recherche effectuées par les utilisateurs. Des analyses statistiques de ces transactions ont été menées aux Etats-Unis sur plusieurs systèmes et dans plusieurs bibliothèques (24). Elles indiquent qu'au moins la moitié des recherches portent sur des sujets, que les utilisateurs ont tendance à répéter les mêmes erreurs, qu'ils restent dans le même type de recherche (sur auteur par exemple) sans explorer d'autres possibilités. L'inconvénient de cette méthode est qu'elle ne nous apprend rien sur les raisons pour lesquelles les utilisateurs font telle ou telle chose. Il est aussi très difficile d'y délimiter les débuts et fins de sessions de recherche pour chaque utilisateur.

Les méthodes classiques d'inter-

views et de questionnaires ont aussi été employées (16), y compris sur des « non-utilisateurs » d'OPAC. Elles nous donnent des indications sur la satisfaction, les préférences et les suggestions des usagers. Là encore on ne récupère aucune information sur leur comportement durant l'interaction au terminal. On questionne les gens une fois leur recherche terminée, et il est difficile de les faire parler en détail de ce qu'ils ont fait.

Ces méthodes sont utiles mais insatisfaisantes si l'on veut évaluer de façon précise et détaillée les interfaces existantes afin de les améliorer. Développer de nouvelles méthodes d'évaluation des systèmes interactifs de consultation par le grand public s'impose si l'on veut qu'ils jouent le rôle crucial qu'ils pourraient avoir, y compris dans les bibliothèques.

RÉFÉRENCES

1. **Gourdiex, Annie**, *L'informatisation des bibliothèques en France : logiciels disponibles*, Marseille, CESIA (Centre d'études des systèmes d'information des administrations), sept. 1984, 295 p., ISBN 2-903665-15-X. (Diffusé par la Documentation française).
2. **Finelli, Françoise, Salomon, Serge**, « Le Catalogage en ligne à la Bibliothèque nationale », *Bulletin d'information de l'ABF* n° 128, 3^e trim. 1985, p. 19-25.
3. **Lalanne, J.** « DOBIS/LIBIS au Ministère des PIT », *Bulletin du Centre de Hautes études internationales d'information documentaire*, n° 11, 3^e trim. 1983, p. 55-62.
4. **Guyot, Brigitte**, « Des minitels dans les bibliothèques ? », *Bull. Bibl. France*, t. 29, n° 3, 1984, p. 240-243.
5. **Barbier-Bouvet, Jean-François**, « Portrait de groupe avec Minitel », *Bull. Bibl. France* t. 29, n° 3, 1984, pp. 230-234.
Bull. Bibl. France, 6.
6. **Belkin, Nick J.**, « Anomalous state of knowledge as a basis for information retrieval », *Canadian journal of information science*, n° 5, 1980, p. 133-143.
7. **Mitev, Nathalie N., Venner, Gillian M., Walker, Stephen**, *Designing an online public access catalogue: Okapi, a catalogue on a local area network*, London, BLRD, 1985, 254 p. (Library and information research reports, no. 39), ISBN 0-7123-3058-5.
8. **Gould, John D, Lewis, Clayton**, « Designing for usability: key principles and what designers think », *Communications of the ACM*, t. 28, n° 3, March 1985, p. 300-311.
9. **Hancock, Micheline**, Travail en cours au Department of information science, The City university, Londres, Grande Bretagne.
10. **Mitev, Nathalie**, « Users and ease of use: online catalogues' raison d'être », *Program*, t. 20, n° 2, April 1986, p. 111-119.
11. **Salmon, Stephen R.**, « Planning for library automation: the origins of MELVYL », *Information technology & libraries*, t. 1, n° 4, Dec. 1982, p. 350-358.
12. **Hildreth, Charles R.**, « Online public access catalogs », *Annual review of information science & technology*, vol. 20, 1985, p. 233-285.
13. **Matthews, Joseph R.**, *Public access to online catalogs*, 2nd ed., New York, Neal-Schuman publishers, 1985, 497 p., (Library automation planning guide series), ISBN 0-918212-89-8.
14. **Manning, Ralph W.**, « L'avenir des catalogues et les catalogues de l'avenir », *Documentation et bibliothèques*, t. 30, n° 1, janv.-mars 1984, p. 7-12.
15. **Kinsella, Janet, ed.**, *Online public access to library files: proceedings of a center for catalogue research conference held at Bath university, England, 3-5 September 1984*, Oxford, England: Elsevier international bulletins, 1985, 191 p., ISBN 0-946395-18-7.
16. **Markey, Karen**, *Online catalog use: results of surveys and focus group interviews in several libraries*, Dublin, Ohio: OCLC, 1983, 264 p., (Final report to the Council on library resources, vol. 2; Research report OCLC/OPR/RR-83/3).
17. **Hildreth, Charles R.**, *Online public access catalogs: the user interface*, Dublin, Ohio, OCLC, 1982, 263 p., (OCLC library, information and computer science series), ISBN 0-933418-34-5.
18. **Mitev, Nathalie**, « The User interface in an online public access catalogue », *Recherche d'information assistée par ordinateur*, Actes du Symposium international RIA 85 organisé par le Centre des Hautes études d'informatique documentaire et l'Institut d'informatique et de mathématiques appliquées de Grenoble, 18-20 mars 1985, Grenoble, IMAG (avec le concours du CNRS), 1985, p. 251-272, ISBN 2-905450-00-2.
19. **Hildreth, Charles R.**, « Pursuing the ideal: generations of online catalogs », *Online catalogs, online reference, converging trends*, Brian Aveney and Brett Butler eds. Proceedings of a library & information technology association preconference institute, 23-24 June 1983, Los Angeles. Chicago, American library association, 1984, p. 31-56, (Library and information technology series, no. 2). ISBN 0-8389-3308-4.
20. **Markey, Karen**, *Subject searching in library catalogs: before and after the introduction of online catalogs*, Dublin, Ohio, OCLC, 1984, 176 p., (OCLC Library, information and computer science series, no. 4). ISBN 0-933418-54-X.
21. **Cochrane, Pauline A.**, « Friendly » catalog forgives user errors: no librarian intervention on a dream online system called « PaperChase », *American libraries*, t. 13, n° 5, May 1982, p. 303-306.
22. **Doszkocs, Tamas E.**, « CITE NLM: natural-language searching in an online catalog », *Information technology & libraries*, t. 2, n° 4, déc. 1983, p. 364-380.
23. **Dowlin, Kenneth E.**, *The Electronic library: the promise and the process*, New York, Neal-Schuman, 1984, 199 p., (Applications in information management and technology series). ISBN 0-918212-75-8.
24. **Tolle, John E.**, Current utilization of online catalogs: transaction log analysis, Dublin, Ohio, OCLC, 1983, 113 p., (Final report to the Council on library resources, Vol. 1; OCLC research report series, no. OCLC/OPR/RR-83/2).