

L'ACCÈS SUJET DANS LES CATALOGUES EN LIGNE

LE CAS DES BIBLIOTHÈQUES UNIVERSITAIRES EN FRANCE

Un bref survol de la littérature montre que l'accès sujet est toujours un domaine de recherche en plein essor. Les deux problèmes majeurs de l'accès sujet, à savoir le taux d'échec élevé (entre 30 et 50 %) et la surcharge d'informations, sont encore d'actualité.

Pour atténuer ces problèmes, divers auteurs ont présenté les caractéristiques d'un catalogue qu'ils désignent sous le nom générique de « catalogue de troisième génération » (1) (2) (3) (4) (5). Cet outil, en plus des fonctionnalités classiques des catalogues de deuxième génération, possède de nouvelles caractéristiques :

1. Nouvelles fonctions d'accès et de navigation :

- interrogation en langage naturel ;
- technique d'aide à la recherche en ligne et à la traduction de la requête ;
- techniques de recherche non booléennes ;
- affichage des documents dans l'ordre de pertinence ;
- *feedback* et reformulation ;
- recherche multilingue ;

- navigation hypertextuelle ;
- intégration des classifications, des listes d'autorités et des mots-clés.

2. Enrichissement du contenu de la base bibliographique :

- enrichissement du contenu des notices MARC (résumé, table de matières, augmentation du nombre de vedettes matières, etc.) ;
- signalement des différents types de documents (sommaire de revues, articles, dossiers, document numérique, etc.).

3. Mise en réseau avec d'autres catalogues : un accès à d'autres bases de données et à des catalogues en ligne distants.

4. Filtrage coopératif : offrir des outils de personnalisation de la recherche, permettre des recherches coopératives entre les différents usagers (novices, usagers distants, bibliothécaires, etc.)

5. Des interfaces plus ergonomiques. Ces caractéristiques sont mises en évidence aussi bien dans de récents prototypes¹ que dans quelques systèmes commerciaux.

Le but de cette étude est de recueillir des données empiriques sur les OPAC (Online Public Access Catalog) des bibliothèques universitaires françaises en général, et plus particulièrement, de faire le point sur l'accès sujet.

1. Gerry McKIERNAN a recensé une partie des projets de catalogues de troisième génération sur son site web CYBERSTACKS : <http://www.public.iastate.edu/~CYBERSTACKS/onion.html>

Analyse du questionnaire

Après avoir effectué un état de l'art concernant l'accès par sujet dans les OPAC, nous avons adapté et enrichi un questionnaire déjà testé en Grande-Bretagne (6). Nous l'avons envoyé aux directeurs des bibliothèques universitaires et des services communs de documentation (SCD) au cours de la période allant de décembre 1996 à mars 1997.

Ce questionnaire est composé de six sections (catalogues manuels, OPAC et fonds documentaires, stratégies d'indexation, stratégies de recherche, affichage de l'index et des renvois, aides disponibles). Sur les 93 bibliothèques universitaires² (BU) recensées, nous avons reçu 84 réponses (47 réponses des SCD, 27 des sections et 10 qui sont inexploitable). Dans notre échantillon, 65 BU possèdent un OPAC.

En plus de l'envoi du questionnaire, nous avons eu accès aux catalogues des BU qui existent sur Internet. Le travail de vérification et de correction des données s'est fait en téléphonant aux responsables des BU. Selon les points du questionnaire, nous avons décidé de regrouper les réponses des SCD avec celles des sections ou de les traiter séparément. Les comptages ne sont donc pas aussi homogènes que nous le voulions, mais la valeur des informations reçues nous permet de donner une vue générale sur l'accès par sujet dans les BU.

Les catalogues

Sur 74 BU, on en dénombre 47 qui conservent encore des supports traditionnels (fiches ou microfiches).

Les catalogues manuels

Actuellement, l'accès manuel ne concerne qu'une partie du fonds documentaire. Ceci impose aux utilisateurs d'effectuer une double

recherche (manuelle et automatisée) pour trouver l'information voulue, ce qui est parfois contraignant. Le premier tableau de l'enquête, sur la typologie des catalogues, propose la répartition suivante :

- catalogue papier : 9
- OPAC : 27
- les deux : 38
- total : 74

**LES DEUX
PROBLÈMES
MAJEURS DE
L'ACCÈS SUJET,
À SAVOIR LE TAUX
D'ÉCHEC ÉLEVÉ
ET LA SURCHARGE
D'INFORMATIONS,
SONT ENCORE
D'ACTUALITÉ**

L'accès manuel se fait d'une façon complémentaire par un catalogue dictionnaire et un catalogue systématique. Les deux modes d'accès les plus fréquents sont l'accès par auteur (47 BU) et l'accès dictionnaire matière (46 BU), comme le révèle le tableau relatif à la typologie des catalogues manuels :

- catalogue dictionnaire auteur : 47
- catalogue dictionnaire matière : 46
- catalogue systématique : 13
- catalogue dictionnaire titre : 6
- autres catalogues : 9

Neuf BU ont, en plus, soit un catalogue topographique, soit un fichier de microfiches OCLC, ou enfin, un catalogue de thèses et mémoires classés par année.

Dans les catalogues manuels, les bibliothécaires emploient deux stratégies d'indexation principales : Rameau (Répertoire d'autorité matière encyclopédique et alphabétique

unifié) et une indexation libre utilisant une liste de descripteurs maison (22 BU). Les langages documentaires utilisés sont donc ventilés de la façon suivante :

- Rameau : 24
- MESH : 7
- autres : 22

Les catalogues en ligne (OPAC)

Soixante-cinq ont fait le choix d'un système intégré de gestion de bibliothèques qui admet un module OPAC. Sept BU n'ont pas un module OPAC séparé, mais un cédérom (CD AUROC). Les systèmes les plus utilisés sont ceux de GEAC (16 BU) et de Dynix (14 BU).

On observe que l'introduction des OPAC dans les BU est récente. On peut relever que plus des deux tiers des BU ont introduit un OPAC ou changé de système depuis 1990.

Qualité du fonds documentaire

Pour l'indexation et la conservation classique des monographies et des thèses, le fonds des BU tend à être de plus en plus multimédia.

Les collections des BU sont donc d'une grande diversité :

- monographies : 65
- thèses et mémoires : 55
- conférences : 42
- périodiques : 42
- images animées (films, vidéo, etc.) : 26
- documents sonores : 23
- microfiches : 21
- logiciels : 7
- documents électroniques : 3
- autres : 8

La catégorie « autres » inclut essentiellement les cartes géologiques (7 établissements).

Si l'OPAC signale la diversité de la collection de la bibliothèque, il ne permet cependant pas d'effectuer des recherches sur des parties de documents. Or ce sont ces parties qui intéressent l'utilisateur. Ainsi, comme le signale Roland Bertrand : « *Les chercheurs s'intéressent prioritairement aux articles des périodiques. Dans un OPAC, ces éléments n'apparaissent pas. On ne trouve que le titre du périodique ou du recueil de conférence.* » (8)

2. Le terme « bibliothèque universitaire » est donc utilisé dans cet article pour désigner aussi bien un SCD qu'une section d'un SCD.

On a observé que la taille des fonds documentaires varie considérablement selon les BU. L'ancienneté, mais aussi l'implémentation géographique (9) expliquent ces disparités. Les notices informatisées se présentent sous différents formats :

- UNIMARC : 28
- LC MARC : 24
- MARC SIBIL : 3
- autres : 7
- sans réponse : 3

Nous remarquons une avancée des BU en matière de normalisation puisqu'elles utilisent en majorité des formats compatibles MARC. La catégorie « autres » inclut les formats suivants : Usmarc (3 BU), Datatrek-marc, Texto ou un format interne (2 BU).

On s'est intéressé au volume des notices informatisées dans un fonds documentaire. Sur les 47 SCD, nous observons qu'une partie importante de ces fonds n'est pas consultable sur un support informatisé.

Il s'agit surtout des BU qui ont un volume important. Les pourcentages de notices informatisées sont ainsi répartis :

- < 25 % : 10
- 25 - 50 % : 9
- 50 - 75 % : 4
- > 75 % : 15
- sans réponse : 7

Malgré l'effort considérable qui a été fait dans le domaine de la rétroconversion, notamment grâce au programme de 1991 (9), les BU sont loin de disposer d'un catalogue entièrement informatisé. Une grande partie des notices sont dépourvues d'indexation matière.

Catalogage et indexation sujet

La majorité des bibliothèques pratique les deux types de catalogage (courant et dérivé), comme l'indique le tableau 7 du questionnaire sur la stratégie d'indexation :

- catalogage dérivé : 63
- catalogage courant : 49
- sans réponse : 2

Le catalogage en local tend de plus en plus à se réduire, et la mise en service du futur système universitaire va sans doute accroître cette tendance.

Le catalogage dérivé à partir des trois réservoirs (BN-Opale, OCLC, SIBIL) est pratiqué sur une large échelle (7). On observe cependant une augmentation du nombre de BU qui utilisent le réservoir BN-Opale.

Les langages documentaires

Les langages documentaires les plus répandus sont Rameau et la classification décimale de Dewey. Nous pouvons constater l'usage fort répandu de la classification Dewey qui a profité de la déréglementation de 1988 (9). La liste de l'ensemble des langages documentaires utilisés est la suivante :

- Rameau : 65
- classification de Dewey : 19
- classification décimale universelle : 4
- classification de la bibliothèque du Congrès : 4
- National Library of Medicine : 2
- classification de Cunningham : 2
- classification maison : 2

Six bibliothèques utilisent plus d'une classification, ce qui pose souvent des problèmes d'incompatibilité liés à l'existence de référentiels hétérogènes. L'utilisation des thésaurus se limite à MESH ou à sa version française le FMESH.

Rameau

L'introduction rapide et généralisée de la liste Rameau est bien perçue par les bibliothécaires, car elle permet d'harmoniser les pratiques d'indexation. Pour la majorité d'entre eux, c'est la garantie que les lecteurs n'auront pas à s'adapter en passant d'une bibliothèque universitaire à une autre. Toutefois, Rameau n'est pas exempt de certains problèmes.

- La plupart des bibliothécaires trouvent l'accès sujet à travers Rameau un peu difficile pour les lecteurs non avertis, qui ont besoin d'une interface permettant de passer facilement du langage naturel à un langage assez structuré du type Rameau.

- Les bibliothécaires déplorent l'absence ou l'insuffisance en nombre de renvois, car « l'utilisation des renvois permet de pallier le manque du langage naturel ».

- L'emploi de plusieurs langages documentaires (par exemple Rameau, MESH) crée des conflits.

- Lorsqu'une partie du fonds est spécialisé, « Rameau est insuffisant, car les vedettes sont très larges, ce qui conduit à affiner l'indexation par des termes libres ».

- Souvent les bibliothécaires regrettent l'insuffisance du nombre de vedettes par notice.

Les clés d'accès

L'informatisation des catalogues a permis d'offrir plus de clés d'accès aux utilisateurs. Cette augmentation donne désormais à l'utilisateur des possibilités nouvelles. Ainsi, 68 % des usagers utilisent des clés d'accès inexistantes dans les catalogues manuels (10).

Pour la majorité des bibliothécaires, la possibilité d'une recherche par mots-clés (mots du titre, mots sujet et mots pris dans toute la notice) permet de contourner la difficulté d'une recherche sujet classique.

Les points d'accès disponibles sont les suivants :

- auteur (nom de personne) : 65
- titre : 61
- mots de titre : 60
- auteur (nom de collectivité) : 58
- mots sujet : 57
- sujet (feuilleter alphabétique) : 55
- indice de classification : 35
- auteur/titre : 31
- autres : 32

La catégorie « autres » comprend les éléments suivants : éditeur, collection, ISBN, numéro d'inventaire, numéro de la notice OCLC, lieu d'édition. Les cinq BU qui se sont dotées de l'OPAC AB6, peuvent offrir à leurs utilisateurs la possibilité d'effectuer des recherches en texte intégral.

Recherches booléennes, recherches avancées

Quarante-huit BU garantissent un accès booléen à leurs catalogues. Comme nous le montre le tableau ci-contre, cela varie selon les points d'accès et selon les opérateurs booléens proposés (ET, OU, SAUF, ET implicite).

Recherches avancées		
points d'accès / opérateurs booléens	troncature	opérateurs d'adjacence
auteur	49	7
sujet	47	9
titre	47	8
mots-clés	34	5
indice de classification	19	3
autres points d'accès	6	2

L'utilisation de la recherche booléenne est difficile pour les usagers non expérimentés, qui ne perçoivent pas la différence existant entre les divers opérateurs. On remarque d'ailleurs une sous-exploitation de ces aides à la recherche. L'opérateur « et » est le plus populaire parmi les usagers. On note, au contraire, une faible utilisation des opérateurs « ou » et « sauf » (seulement 2 % de cas) (10).

Pour pallier le problème de la formulation des requêtes, diverses solutions issues de l'intelligence artificielle (système expert, logique floue, recherche probabiliste, langage naturel, etc.) sont proposées (11) (12) (13).

Rares sont les OPAC qui intègrent ces techniques avancées ; en effet, un seul des catalogues de notre échantillon est doté d'une recherche phonétique, et seule la troncature est vraiment présente dans les OPAC.

Les arbres de décisions

Comme la majorité des OPAC existants incorporent différentes stratégies d'interrogation par sujet (sujet alphabétique, mots sujet, utilisation des opérateurs booléens, affichage des renvois, etc.), il est nécessaire de développer un mécanisme permettant d'orienter le type de recherche selon les caractéristiques des requêtes des utilisateurs : c'est le mécanisme des arbres de décisions. L'hypothèse sous-jacente est que chacune des stratégies permet de retrouver des notices bibliographiques différentes. Les arbres de décisions sont implantés dans des prototypes comme ASTUTE³ (14) ou OKAPI (15) et dans un catalogue opérationnel comme COPAC⁴ (16).

Joan Cherry (17) a testé l'efficacité de l'une de ces stratégies en effectuant des recherches par mots du titre et par mots du sujet chaque fois qu'un accès sujet (feuilletage alphabétique ou par phrases) échoue. Elle a trouvé que le rappel augmente d'une façon significative.

Cette caractéristique n'est pas implémentée dans la majorité des OPAC. Seuls deux systèmes permettent de suggérer aux utilisateurs d'effectuer une recherche par mots du sujet ou par mots du titre, lorsque la recherche par sujet échoue. Sinon, la majorité des OPAC dans les BU termine la session par des messages du genre : « pas de réponse à votre demande », « zéro réponse ».

Navigation

Un autre mode de recherche apprécié par les utilisateurs est le butinage de l'index. L'incorporation des liens (synonymie, hiérarchie, association) existant dans cette liste d'autorité a souvent été présentée par les bibliothécaires comme l'une des solutions au problème de l'accès sujet dans les

**IL EST NÉCESSAIRE
DE DÉVELOPPER
UN MÉCANISME
PERMETTANT
D'ORIENTER
LE TYPE DE
RECHERCHE
SELON LES
CARACTÉRISTIQUES
DES REQUÊTES
DES UTILISATEURS :
C'EST LE MÉCANISME
DES ARBRES
DE DÉCISIONS**

Recherche booléenne				
points d'accès / opérateurs booléens	ET	OU	SAUF	ET implicite
auteur	31	33	25	26
sujet	31	32	24	30
titre	30	32	25	31
mots-clés	31	29	25	25
indice de classification	14	16	14	10
autres points d'accès	8	7	5	6

catalogues. Cinquante-cinq BU permettent un affichage de l'index, mais la plupart des catalogues actuels n'exploitent pas les structures de renvois de la liste Rameau. L'affichage des renvois est le suivant :

- voir : 43
- voir aussi : 25
- TG, TS, TA : 0
- sans réponse : 1

3. ASTUTE : A Search Tree Underlying The Experiment.

4. COPAC : Consortium of University Research Libraries OPAC.

**BEUCOUP DE
BIBLIOTHÉCAIRES
CONSIDÈRENT
QUE LE CONTENU
DE LA FORMATION
DOIT, LUI AUSSI,
ÉVOLUER,
IL NE DOIT
PLUS S'ARTICULER
SEULEMENT
SUR LES ASPECTS
BIBLIOTHÉCO-
NOMIQUES ...**

Pour les bibliothécaires, les raisons qui expliquent l'absence des renvois sont :

- l'incapacité du système informatique à gérer les renvois ;
- le manque de personnels ;
- les différents modes de recherche existants qui rendent l'affichage de renvois redondant ;
- des raisons budgétaires.

L'absence des relations dans les OPAC n'est pas due uniquement à des problèmes informatiques ou humains, mais à la complexité du langage Rameau (18).

Une façon simple d'incorporer des possibilités de navigation dans l'OPAC consiste à habiller l'interface par une couche web.

Contrairement aux OPAC classiques, les WWW-OPAC permettent maintenant une navigation non linéaire dans la base bibliographique ; ils favorisent ainsi la construction dynamique et exploratoire des recherches d'information. Disposer d'une interface Web permettra aux usagers de ne plus se limiter aux stratégies de recherche faibles que sont l'essai-erreur ou le feuilletage alphabétique,

mais au contraire, de mettre en œuvre certaines stratégies des professionnels, comme la possibilité de relancer une recherche avec les termes d'indexation ou les clés d'accès d'une référence déjà trouvée : c'est une forme de reformulation itérative dirigée par l'utilisateur final. Nous appelons cette stratégie : Browsing Relevance Feedback (BRF). Nos analyses ont montré que la stratégie BRF est souvent employée par les usagers (19). Elle permet de remédier à l'absence des techniques de reformulation dans les catalogues en ligne.

**Mode de diffusion
et formation des usagers**

Le travail en réseau est une pratique importante et courante de la communauté universitaire. Un certain nombre d'utilisateurs des BU (notamment les chercheurs et étudiants de 3^e cycle) ne peuvent se contenter d'un accès local. L'accès distant à d'autres ressources est devenu primordial pour eux. Ils peuvent désormais interroger des milliers de catalogues accessibles sur Internet à partir de leurs postes de travail. La diffusion du protocole de recherche et de repérage Z39.50 va sans doute accentuer cette tendance.

Environ la moitié des BU offre un accès distant à leurs ressources documentaires par le biais du Web.

La plupart des bibliothécaires estiment que la formation des utilisateurs et la disponibilité des bibliothécaires sont des moyens de contourner la difficulté d'accès à un OPAC, car les aides en ligne ne sont pas suffisantes. Les aides disponibles sont ainsi ventilées :

- disponibilité des bibliothécaires : 62
- aide en ligne sur écran : 51
- aide-mémoire, brochure : 35
- session de formation : 32
- panneau d'affichage : 28
- didacticiels : 2
- sans réponse : 3

Beaucoup de bibliothécaires considèrent que le contenu de la formation doit, lui aussi, évoluer. Il ne doit plus s'articuler seulement sur les aspects bibliothéconomiques, mais prendre

de plus en plus en compte des aspects techniques de téléchargement, d'impression, de formatage et d'accès à distance.

Un effort soutenu

Dès le début des années 90, un effort considérable et soutenu a été réalisé pour l'informatisation des bibliothèques universitaires. Les résultats de cette étude montrent que cet effort est maintenu depuis sept ans, aussi bien au niveau de la rétroconversion que de celui de la normalisation. D'après les réponses reçues, la majorité des catalogues de notre échantillon est de deuxième génération. La disponibilité sur le marché de systèmes commerciaux basés sur le modèle non booléen, la diffusion de la norme Z39.50 et l'introduction du web sont trois facteurs qui peuvent améliorer la recherche dans les catalogues en ligne.

Nous sommes d'accord avec Carolyn Frost (2), lorsqu'elle signale que toute évolution des OPAC doit répondre aux trois objectifs de Cutter⁵ (20) :

1. *Rechercher* : un catalogue doit permettre de trouver un livre dont on connaît l'auteur, le titre ou le sujet.
2. *Regrouper* : il doit pouvoir indiquer quels ouvrages de tel auteur la

**... MAIS PRENDRE
DE PLUS EN PLUS
EN COMPTE
DES ASPECTS
TECHNIQUES DE
TÉLÉCHARGEMENT,
D'IMPRESSION,
DE FORMATAGE
ET D'ACCÈS
À DISTANCE**

bibliothèque possède, quels ouvrages elle possède sur un certain auteur ou sur un certain type de littérature.

3. *Assister* : il doit aider à choisir un ouvrage d'après son édition ou son caractère (littérature ou documentaire).

Si les OPAC répondent au premier objectif, rares sont ceux qui répondent aux deux derniers. Cette transformation des catalogues ne sera effective qu'en dépassant la rigidité⁶ du format MARC – SGML est une bonne alternative (13) – et en enrichissant le contenu des notices bibliographiques.

Mars 1998

BIBLIOGRAPHIE

1. **Beheshti, Jamshid.** – « The evolving OPAC ». – *Cataloging and Classification Quarterly*, vol. 24 (1-2), 1997, p. 163-185.
2. **Frost, Carolyn.** – « Next-generation online public access catalogs ». – *Advances in Library Automation and Networking*, vol. 5 (1994), p. 1-41.
3. **Hildreth, Charles.** – « An evaluation of structured navigation for subject searching in online catalogs ». – *PhD dissertation*, Departement of Information Science, London, 1993.
4. **Matthews, Joseph.** – « Time for new OPAC initiatives : an overview of landmarks in the literature and introduction to wordfocus ». *Library Hi Tech*, vol. 15 (1-2), 1997, p. 111-122.
5. **Khoo, C. S ; Poo, D. C.** – « Subject access in online catalogs ». – *Encyclopaedia of Library and Information Science*, vol. 60 (1997), p. 324-340.
6. **Crawford, John.** – « A survey of subject access to academic library catalogs in Great Britain ». – *Journal of Librarianship and Information Science*, vol. 25, 1993, p. 85-93.
7. **Thoumieux, Roger.** – « Les bibliothèques universitaires ». – *Bulletin des Bibliothèques de France*, t. 37, n° 2, 1993, p. 62-70.
8. **Roland, Bertrand.** – « Le catalogue, les bibliothèques et la modernité ». – *Bulletin des Bibliothèques de France*, t. 36, n° 4, 1991, p. 295-316.
9. **Renoult, Daniel.** – *Les Bibliothèques dans l'université*. – Paris : Éd. du Cercle de la librairie, 1994.
10. **Peters, Thomas.** – « The history and development of transaction log analysis ». – *Library Hi Tech*, vol. 42, 1993, p. 41-66.
11. **Le Loarer, Pierre.** – « OPAC : Opaque or Open, Public, Accessible and Cooperative ? Some developments in natural language processing ». – *Program*, vol. 27, 1993, p. 251-268.
12. **Andreasen, Troels ; Schomacker, Tommy.** – « DanBib : a union catalogue applied for user friendly flexible querying ». – 63rd IFLA conference, Copenhagen, 1997.
13. **Larson, Ray [et al.].** – « Cheshire II : designing a next generation online catalog ». – *Journal of the American Society for Information Science*, vol. 47 (7) 1996, p. 555-567.
14. **Drabenstott, Karen M. ; Vizin-Goetz, Diane.** – *Using Subject Heading for Online Retrieval : Theory, Practice and Potential*. – San Diego : Academic Press, 1994.
15. **Walkers, S.** – « OKAPI : evaluating and enhancing an experimental online catalog ». – *Library Trends*, vol. 35 (4), 1987, p. 631-645.
16. **Cousins, Shirley Anne.** – « COPAC : the new national OPAC service based on the CURL databases ». – *Program*, vol. 31, 1997, p. 1-23.
17. **Cherry, Joan.** – « Improving subject access in OPAC : an exploratory study of conversion of users' queries ». – *The Journal of Academic Librarianship*, vol. 18 (2), 1992, p. 95-99.
18. **Ihadjadene, Majid.** – « L'organisation des connaissances dans les bibliothèques universitaires en France ». – Première conférence du chapitre français de l'ISKO, Lille 1997.
19. **Ihadjadene, Majid.** – « La recherche et la navigation dans un WWW-OPAC : le cas des utilisateurs distants ». – 26th Annual CAIS Conference, Ottawa, Canada, juin 1998 (à paraître).
20. **Mitev, Nathalie ; Hildreth, Charles.** – « Les catalogues interactifs en Grande-Bretagne et aux États-Unis ». – *Bulletin des Bibliothèques de France*, t. 34, n° 1, 1989, p. 23-47.

5. Charles Ami CUTTER est l'une des premières personnes à définir les objectifs d'un catalogue, dès le début du siècle, dans *Rules for a Dictionary Catalog*, Whashington, 1904. Il fait référence aux États-Unis.

6. L'une des limites du format MARC est qu'il ne traite pas bien les différentes relations qui existent entre les œuvres bibliographiques. Bradford-OPAC est un bon exemple de catalogue qui prend en compte ces relations. Il est accessible sur : <http://www.comp.brad.ac.uk/research/database/bopac2.html>