

Les Catalogues interactifs

Madjid Ihadjadene

*Ecole Nationale Supérieure des Sciences
de l'Information et des Bibliothèques (ENSIB)
17-21 bd du 11 nov. 1918 Villeurbanne 96100 France
Ihadjde@ensib.fr*

Définition :

Un catalogue interactif ou OPAC (Online Public Access Catalogue) est un système informatique permettant l'interrogation à distance de catalogues automatisés de bibliothèques directement par l'utilisateur final. Pour rappel un catalogue de bibliothèque est un ensemble de notices catalographiques des documents composant un fonds documentaire. Ces notices sont rédigées selon des principes normalisés et classées afin de faciliter les recherches des usages.

Les catalogues interactifs ont été mis en œuvre dès la fin des années 60 dans un certain nombre de bibliothèques, notamment dans les pays anglo-saxons. D'un point de vue conceptuel, un catalogue interactif est un système de recherche d'informations (SRI) qui a les caractéristiques suivantes:

- Une interface utilisateur qui gère le dialogue entre l'utilisateur et le système.
- Une fonction d'appariement et d'interrogation
- Une base de données constituée des notices MARC (Machine Readable Catalog) et d'un ensemble d'index.

Cependant, il existe une différence significative entre les SRI et les OPACS et ce du point de vue de:

◆ **Population étudiée :**

- Dans les SRI, le public concerné est spécialisé, c'est celui des professionnels de l'information, alors que le public des catalogues est hétéroclite.
- Les SRI sont utilisés par des intermédiaires en information (documentaliste, bibliothécaires, ... etc) alors que les catalogues interactifs sont consultables directement, sans intermédiaire professionnel, par des usagers finals, non formés et inexpérimentés.
- Les stratégies de recherche des intermédiaires tendent à être formelles et planifiées alors que les usagers consultent les catalogues d'une façon formelle.

◆ **Base de données :**

- Le contenu des bases de données d'un SRI est spécialisé et compose souvent d'articles, alors que dans les bibliothèques, le fonds documentaire est encyclopédique et contient des monographies.

- Les enregistrements présents dans les catalogues interactifs sont souvent dépourvus de résumés ou de tables de matière dans les OPACS, le format des données est normalisé (souvent le format MARC) alors que dans les SRI, ce n'est pas le cas.
- ◆ **Vocabulaires contrôlés :**
- Dans les SRI, on emploie souvent des thésauri pour l'indexation et la recherche d'information alors que dans les catalogues, on utilise soit des listes d'autorité.
- ◆ **Coût:**
- Les coûts monétaires peuvent être un élément important dans une consultation d'un SRI alors que la consultation d'un catalogue est libre et gratuite.

Ces paramètres sont résumés ci-dessous:

	OPACs	SRI
Population concernée	- public hétéroclite et non formé	- public spécialisé et formé aux techniques documentaires
Médiation	Utilisation de l'OPAC sans intermédiaire.	- interrogation des SRI avec l'aide d'un intermédiaire (bibliothécaire, documentaliste, archiviste)
Base de donnée	- le fonds documentaire est encyclopédique - contient des monographies - les enregistrements d'une notice bibliographiques sont dépourvus de résumés ou de tables de matières - utilisation d'un format, MARC en général	- le contenu de la base est spécialisé - composée d'article - les notices documentaires contiennent des résumés - absence de format
Vocabulaires contrôlés	- emploi d'une liste d'autorité et de classifications encyclopédiques (classification décimale de Dewey, classification décimale universelle ... etc)	- emploi d'un thesaurus
Fonction d'appariement	- système booléen	- système booléen mais aussi probabiliste et vectorielle
Coût	- consultation libre et gratuite	- consultation payante

Tableau : comparaison entre un OPAC et un SRI

II. Les Objectifs d'un catalogue :

Charles Ami Cutter (1) est sans doute le premier au 19^{ème} siècle à énoncer les objectifs d'un catalogue:

- 1) Rechercher (finding function): un catalogue doit permettre de trouver un livre dont on connaît l'auteur, le titre ou le sujet.
- 2) Regrouper (collective function): il doit pouvoir montrer ce que possède la bibliothèque sur un certain auteur ou sur un certain type de littérature.
- 3) Assister (assisting function): il doit aider à choisir un ouvrage d'après son édition ou son caractère (littérature ou documentaire).

A la conférence² internationale sur les principes de catalogage de Paris en 1961 (2); seul le premier objectif a été entièrement approuvé, les deux autres ne le sont que partiellement.

III. Les études sur les catalogues interactifs:

Hidreth (3) distingue trois générations d'OPACs qui correspondent à différents principes d'automatisation de l'accès direct au catalogue.

- ◆ La première génération se caractérise par des outils d'accès très rudimentaires. Ils stimulent la recherche d'information dans les catalogues traditionnels (microfiches). Ils n'offraient que la possibilité d'un appariement exact par phrases d'où leur efficacité pour retrouver des données exactes (pour chercher des ouvrages déjà connus). Par contre, l'utilisateur a des difficultés à trouver une réponse s'il ne connaît pas à l'avance l'ordre exact de la phrase (nom de l'auteur, le titre exacte, ... etc). En terme de dialogue homme-machine, ces catalogues offrent en général une approche par menus.

Ces catalogues ont les défauts suivants:

- le manque de points d'accès comme la recherche sujet
- une interface peu ergonomique: ils n'offraient pas de messages d'aides, ils utilisaient un langage spécialisé (celui des bibliothécaires) dans le dialogue avec l'utilisateur.
- les possibilités de navigation sont limitées.
- Pas de recherche par mots clés.

Les premières études effectuées sur ce type de catalogue montrent que la consultation de ce type de catalogue n'est pas plus performante que celle d'un catalogue manuel, certains auteurs estiment même qu'ils sont inférieurs aux catalogues papiers (8).

²Ces principes connus sous le nom de "principe de Paris" sont à l'origine des règles de catalogage dans plusieurs pays, notamment les AACR en Etats-Unis.

Les catalogues de deuxième génération ont intégré les fonctionnalités des SRI comme la recherche booléenne (recherche post-cordonnée), multiplication des critères de recherche, l'utilisation de la troncature et des opérateurs de proximité, les possibilités de limite et de tri des recherches, l'accès par un vocabulaire contrôlé, l'affichage des vedettes matières et des renvois "voir" et "voir aussi".

Avec la recherche post-coordonnée (par mots) l'utilisateur n'a plus besoin de connaître des phrases exactes.

- ◆ Il y a eu aussi des améliorations au niveau de l'interface et du dialogue comme l'introduction des interfaces graphiques, la possibilité d'interrogation par commande et par menu, la disponibilité de plusieurs formats d'affichages, l'affichage de l'historique de recherche et une amélioration du contenu des aides en lignes et des messages d'erreurs.

L'ensemble des éditeurs de logiciels documentaires offre ce type de système comme GEAC³, DRA⁴, EVER⁵ ou VTLS⁶.

La majorité des catalogues interactifs existants en France (4) et dans les pays anglo-saxons sont de deuxième génération (5).

Dès le début des années 80, un ensemble d'études (6) (7) (8) ont été effectuées sur l'utilisation de ces catalogues par les usagers dont voici les principales conclusions.

- Approximativement 67% des sessions de recherches concernent seulement un seul mode de recherche (sujet, auteur, titre, ... etc).
- Lorsqu'on leur offre le choix entre l'accès en langage de commande, et par menu, seuls 10% des usagers préfèrent le mode d'accès par commande, sinon la majorité d'entre eux optent pour un accès par menu.
- Les recherches par sujet sont les plus fréquentes dans les OPACs et aussi ce sont elles qui aboutissent aux taux d'échecs les plus élevés. Les usagers ont du mal à trouver les termes adéquats et à formuler leurs requêtes. Markey (6) trouve que presque 45% des accès sujet aboutissent à des échecs. Divers facteurs expliquent cette inadéquation entre des systèmes informatiques et les besoins des utilisateurs dont l'un est lié à la complexité de l'indexation matière.
- Un autre problème concernant la recherche dans les catalogues est celui de la surcharge d'informations lorsqu'une équation aboutit à l'affichage des plusieurs références.
- Ils effectuent souvent des erreurs de type mécanique (erreurs d'utilisation des touches de clavier, ou de commandes) mais aussi des erreurs de type syntaxiques et typographique. Borgamn montre par exemple que 13% des commandes utilisées contiennent des erreurs typographiques ou logiques (7).

³[Http://www.geac.com](http://www.geac.com)

⁴[Http://www.dra.com](http://www.dra.com)

⁵[Http://www.ever.fr](http://www.ever.fr)

⁶[Http://www.vtls.com](http://www.vtls.com)

- Ils ont des difficultés pour sélectionner les bases de données à interroger.
- Environ un tiers des usagers consulte les aides, surtout au début de leurs recherches.
- Les recherches des usagers sont très simples, souvent ils utilisent un ou deux mots dans leurs requêtes, ils emploient rarement les opérateurs booléens, ou la troncature.
- Contrairement aux spécialistes, ils n'emploient pas de stratégies de recherche.

- 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 (3) (6) (10) (11).

Ces outils, en plus des fonctionnalités classiques des catalogues de deuxième génération, possèdent de nouvelles caractéristiques:

Nouvelles fonction 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
- Feed-back et reformulation
- Recherche multilingue
- Navigation hypertextuelle
- Intégration des classifications, des listes d'autorités et des mots-clés

Enrichissement du contenu de la base bibliographique:

- Enrichissement du contenu des notices MARC (résumé, table de matière, augmenter le nombre de vedettes matières ... etc).
- Signalement des différents types de documents (sommaire de revues, articles, dossiers, document numérique ...etc).

Mise en réseau avec d'autres catalogues :

Un accès à d'autres bases de données et à des catalogues en lignes distants. La généralisation de l'architecture client/serveur basée sur la norme Z39.50 va sans doute accroître cette tendance.

Filtrage coopératif :

Offrir des outils de personnalisation de la recherche, permettre des recherches coopératives entre les différents usagers, permettre des recherches coopératives entre les différents usagers grâce à collecticiels et offrir des outils de recherche personnelle comme les annotations.

⁷Gery McKirnan a recensé une partie des projets de catalogues de troisième génération dans son site web CYBERSTACKS:"<http://www.public.iastate.edu/CYBERSTACKS/union.html>.

Des interfaces plus ergonomiques :

Ces caractéristiques sont mise en évidence aussi bien dans de récents prototypes⁷ que dans quelques systèmes commerciaux.

L'un des développements les plus importants de ces années est l'intégration des techniques hypertextuelles dans les catalogues interactifs.

IV. Les hypercatalogues :

Hjerpe (12) constate que l'un des paradoxes de la recherche d'information vient de ce que l'utilisateur doit décrire l'information qu'il n'a pas "the need to describe that which you do not know in order to find it".

Darvin et Belkin (cités par Hildreth), ayant étudié les besoins d'informations des usagers, montrent que ces derniers ne se servent d'un SRI (OPAC) que lorsqu'ils ont reconnu qu'ils ne savent pas quelque chose, qu'il y a une brèche dans leurs connaissances d'où leurs difficultés à décrire ce qu'ils ne connaissent pas: c'est pour cette raison, que plusieurs chercheurs ont suggéré d'incorporer les possibilités navigationnelles dans les catalogues interactifs: les hypercatalogues.

Dans le passé, seuls quelques prototypes avaient cette fonctionnalité (navigation) comme: HYPERCAT (12), SLC (13), WHIZRD (14) et PACE (10).

Une façon simple d'incorporer des possibilités de navigation dans un catalogue interactif consiste à habiller l'interface par une couche WWW constituant ainsi un hypercatalogue.

Contrairement aux OPACs classiques, les WWW-OPACx 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 WWW permettra aux usagers de ne plus se limiter aux stratégies de 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).

L'analyse de 200 sessions de recherches nous a montré que près de 38% des usagers distants emploient cette stratégie (15).

La stratégie Browsing Relevance Feedback permet ainsi de remédier à l'absence des techniques de reformulation dans les catalogues interactifs.

Afin d'améliorer la recherche et la navigation dans ces hypercatalogues, nous avons suggéré d'améliorer la conception de ces systèmes selon deux niveaux (fig.1)

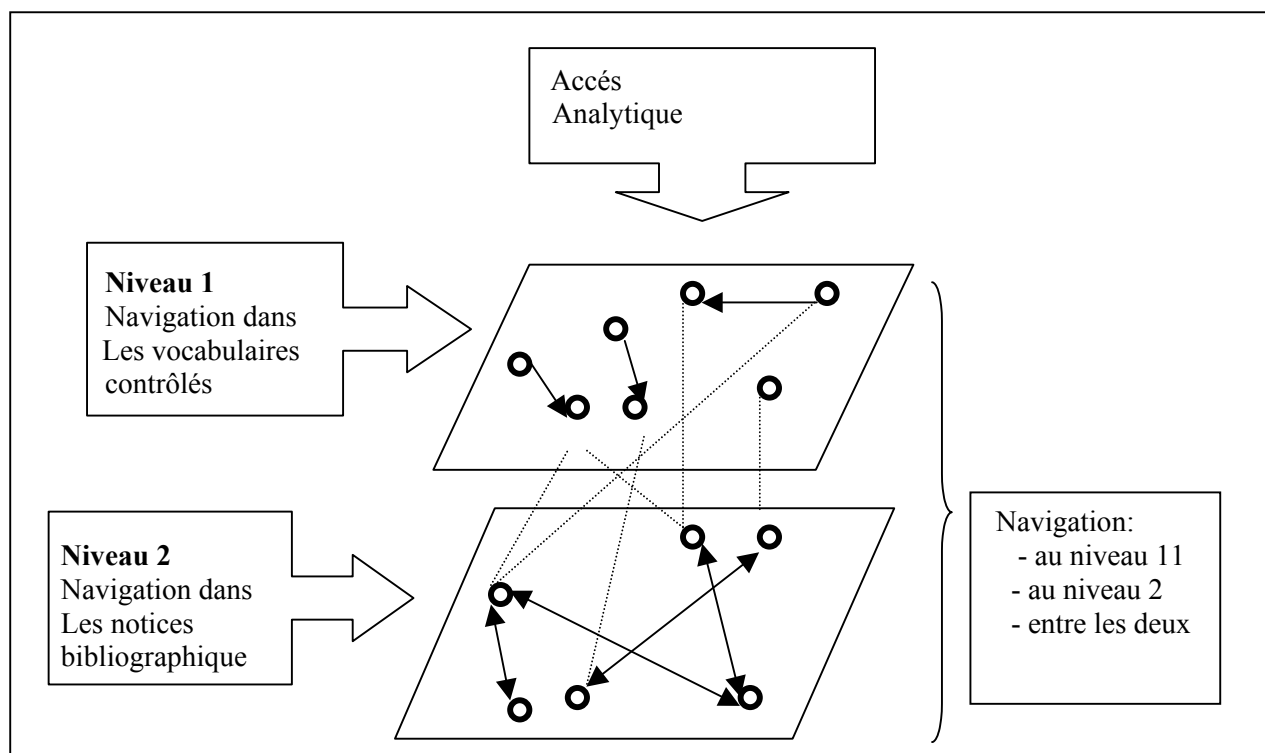


Figure 1. Schéma d'un hypercatalogue

Dans le premier niveau, il s'agit de représenter les vocabulaires contrôlés (classifications, thesaurus ou la liste d'autorité) sous forme d'un hypertexte (16). Dans le cas d'un thesaurus, les nœuds représentent descripteurs et les liens représentent les relations suivantes:

- TG: termes générique
- TA: termes associés
- TS: termes spécifiques

Pour un usager novice qui ne possède pas les termes du domaine qui lui permettent d'exprimer son besoin d'information, représenter graphiquement ou sous forme d'un hypertexte le vocabulaire.

Dans le second niveau, il s'agit d'améliorer la définition des relations qui existent entre les œuvres bibliographiques (17) (18) (19) (20). En effet, rares sont les catalogues (et les SRI) qui permettent à l'utilisateur de naviguer dans les notices bibliographiques. Une taxonomie de ces relations a été définie par Tillet (17), elle regroupe:

- les relations d'équivalence qui relient une œuvre qui existent sur différents supports.
- les relations dérivées qui relient une œuvre et les différentes versions qui en découlent.
- les relations descriptives qui relient une œuvre et ses critiques et commentaires.
- les relations tout-partie qui unissent un élément constitutif d'une unité bibliographique et cette œuvre.
- les relations d'accompagnement qui lient une unité bibliographique et un document annexe.
- les relations séquentielles qui lient des œuvres bibliographiques qui se suivent mais ne sont pas dérivées.
- Les relations de caractéristiques communes qui relient les œuvres bibliographiques qui ont des éléments en commun (auteur, sujet, indice Dewey, ...etc).

Tillet montre que presque 75% de la base MARC du catalogue de la bibliothèque du congrès, admettent un ou plusieurs type de ces relations. Smirglia (20) estime que près de 49.9% de son corpus ont des relations dérivées alors que Vellucci (18) montre que près de 97% des notices d'une base (musicale) ont au moins un lien avec une autre notice.

Conclusion :

L'un des facteurs majeurs dans la généralisation de l'introduction des catalogues interactifs dans les bibliothèques (en France et dans les pays anglo-saxons) est l'adoption par celles-ci des normes nationales et internationales de description et de présentation des documents afin d'améliorer l'interopérabilité des systèmes. Dès le début des années 90, divers facteurs ont influencé le développement des catalogues en ligne (OPACs) dont:

- Le passage d'une architecture centralisée à une architecture client/serveur.
- L'introduction de l'Internet en général et de WWW en particulier.
- La possibilité aux usagers d'accéder à des catalogues distants par une même interface notamment par la diffusion de la norme Z39.50
- Le fonds documentaire des bibliothèques devient de plus en plus multimédia et incluant des documents électroniques.*

Ces facteurs vont sans doute influencer la recherche autour des OPACx aussi bien au niveau quantitative et bien sur qualitative (méthodologiques).

Références Bibliographiques

- 1). **Cutter Charles.**-Rules for a dictionary catalog - 4th edition, Washington DC, 1904
- 2). **FIAB** - Conférence Internationale sur les principes de catalogage- Bulletin des bibliothèques de France, 1961
- 3). **Hildreth Charles.**- An evaluation of structured navigation for subject searching in Online catalog- 1993, PHD dissertation, departement of information science, London
- 4). **Ihadjadene Madjid.**- L'accès sujet dans les catalogues en lignes: le cas des bibliothèques universitaires en France-In Bulletin des Bibliothèques de France, vol.34 n°4, 1998.PP.104-110.
- 5). **Crawford Jonh.** "A survey of subject access to academic library catalogues in Great Britain- In Journal of librarianship and information science- vol. 25, 1993. PP.85-93
- 6). **Karen M Drabenstott; Diane Vizin-Goetz.**- Using subject heading for online retrieval.theory, prectice and potencial- Edition du San Diego. Academic Press, 1994
Borgman
- 8). **Mitev Nathalie; Hildreth Charles.**- "Les catalogues interactifs en Grande-Bretagne et aux Etats-Unis"- Bulletin des Bibliothèques de France - vol. 34, n°1, 1991. PP.23-47
- 9). **Peters Thomas.**- The history and development of transaction log analysis. In Library Hi Tech - vol 42, 1993. PP. 41-66
- 10). **Bheshti; Jamshid.**- OPAC.- in Cataloging and Classification Quarterly, vol. 24, n°1, 1997, PP.153-185
- 11). **Frost Carolyn.**- Next-generation online public access catalogs. In Advances in library automation and Networking, vol. 5 (1994), PP. 1-41
- 12). **Hjerppe R.**- Projet hypercatalog: visions and preliminary conceptions of an extended and enhaced catalog- In actes du 6^{ème} congrès de l'Intelligent Information Systems for the Information Society. Frascati Italy 1985.
- 13). **Horgman C et al.**- The science library catalog projet - In the proceeding of the 54th ASIS annual meeting, 1991. PP. 162-169
- 14). **Nelson M.**- The design of a hypertext interface for information retrieval- In the Canadian journal of information science, vol. 16, n°2, PP. 1-12.

15). Ihadjadene Madjid.- "Searching and surfing a WWW-OPAC: the ease of remote users - In 26th annual Canadian Association of Information Science conference, Ottawa, Canada, 3-5 June 1998 (à paraître).

16). Ihadjadene Madjid.- Hypertexte et vocabulaires contrôlés - In Mosaiques sous la direction de C. Dentzer, Edition de l'ADBS, 1997

17). Tillet B.- Bibliographic relationships: towards a conceptual structure of information used in cataloguing - 1987 PHD thesis University of California.

18). Velluci S.- Bibliographic relationships - In International conference on the principles and the future development of AACR2, Canada, 1997

19). Fattahi R.- The relevance of cataloguing principles to the online environment, 1996, PHD thesis. University of New South Wales.

20). Smiraglia.- Authority control and the extent of derivative bibliographic relationships, 1992. PHD thesis. University of Chicago.