

Recherches multibase, de nouveaux outils pour accroître l'autonomie des usagers

Des usagers submergés par la multiplication des ressources électroniques

Avec le développement des sites Web gratuits ou payants, bibliothèques et centres de documentation voient se multiplier les sources d'information susceptibles de compléter les bases internes. Ces sites Web disposent le plus souvent de leurs propres bases de références aux logiques de recherche particulières. En l'état, cette diversité des ressources n'est susceptible de répondre qu'aux besoins des documentalistes chevronnés. Pour tirer le meilleur profit de ces sources d'informations, il faut non seulement connaître le contenu des bases (information généralement occultée par leurs éditeurs) mais également le traitement documentaire qui a guidé leur constitution (absence d'indexation en raison d'une recherche de type texte intégral, utilisation de thésaurus, indexation des sujets...). L'utilisateur néophyte est affecté d'un double handicap : il ne sait pas repérer la base la plus pertinente et lorsqu'il l'identifie il ne sait comment formuler sa requête (doit-il utiliser des termes précis ? des termes plus génériques ? faut-il essayer avec des formes anglaises ou les formes françaises donneront-elles le même résultat ? faut-il essayer toutes les formes d'un mot ? tous les mots d'une même famille ?).

L'accessibilité du Web n'est pas un leurre mais elle se réduit à la généralisation des outils de consultation (tout ordinateur doté d'un navigateur) alors que les stratégies de recherche d'information se compliquent de manière exponentielle avec l'augmentation continue du nombre de sites. Finalement, alors que l'on pouvait craindre une marginalisation des documentalistes ou des bibliothécaires dans l'accès à l'information, la multiplication des ressources électroniques renforce le besoin d'un médiateur professionnel. L'usager en est-il convaincu ? Le documentaliste lui-même est-il d'accord avec ce constat ou se laisse-t-il séduire comme l'usager par la facilité de manipulation du navigateur et des moteurs de recherche ? Rien n'est moins sûr. Pourtant ni l'usager, ni le documentaliste n'ont vraiment le choix. L'information a un coût et les frais de personnel en constituent le premier poste de dépense. La recherche d'une plus grande efficacité passe par le développement de l'autonomie de l'usager : le plus grand nombre doit arriver à se débrouiller seul pour que le professionnel puisse se consacrer à la population plus marginale des usagers ayant des recherches complexes à effectuer ou rencontrant de véritables difficultés d'accès à l'information. C'est une problématique commune aux centres de documentation d'entreprise (qui sont certainement guidés par la recherche d'une maîtrise des coûts) comme aux bibliothèques publiques (qui doivent libérer du personnel qualifié pour rendre la culture ou l'information accessible à tous).

Quatre solutions techniques pour faciliter l'accès du public à des informations dispersées

Deux de ces solutions reposent sur un enrichissement des données ou des fonctionnalités du catalogue alors que les deux autres proposent des outils complémentaires.

Mieux articuler catalogue et sites Web en stockant l'adresse des sites en zone 856 de la notice Unimarc

C'est l'une des premières techniques qu'ont utilisées les bibliothèques pour effacer la ligne de partage entre la documentation papier et les ressources électroniques. Dans la notice de l'Encyclopaediae Universalis papier, il est fort simple de stocker l'adresse du site Web donnant accès à la version électronique du document. L'ergonomie des opacs aidant, l'usager est incité à se tourner vers cette base en ligne en cliquant simplement sur une icône prometteuse. Si cette technique est relativement adaptée au cas cité, elle atteint rapidement ses limites. Ainsi pour faciliter l'accès à un périodique en ligne, il faudra saisir l'adresse du périodique en zone 856. Comme la plupart du temps, cet accès est payant et que la bibliothèque a recours aux services d'un groupeur, il va falloir saisir une adresse relativement complexe comprenant le nom du site Web, la désignation explicite du périodique et éventuellement des informations d'accès (mot de passe ou équivalent). La construction de telles adresses est lourde (il faut du temps pour construire chaque adresse) et complexe (il faut maîtriser la syntaxe de cette adresse, ce qui nécessite une connaissance approfondie de chaque site Web sollicité). Si la bibliothèque dispose de plusieurs centaines de titres de revues, si elle recourt aux services de plusieurs groupeurs et si l'on tient compte du caractère relativement éphémère de ces adresses (les évolutions d'un site Web aboutissent à des changements de forme de l'adresse), on mesure mieux la charge de ce travail qui semblera très rapidement des plus ingrats.

Utiliser les moteurs de recherche du Web

C'est une solution dont le domaine d'emploi mérite d'être précisé. Faussement simple, elle est à la fois partielle, trompeuse et difficile à maîtriser. Si l'utilisateur est assuré de toujours trouver quelques milliers de pages répondant à sa requête, il passe à côté de l'énorme masse d'informations structurées accessibles au travers de serveurs générant des pages Web dynamiques. La plupart des moteurs de recherche ne peuvent (et même s'ils le peuvent, ils n'en ont généralement pas le droit) référencer et indexer les pages des catalogues de bibliothèque ou des banques de données. Ils traitent principalement les pages Web statiques. La tromperie tient à la fois à ce traitement partiel des informations et au bruit colossal que génèrent ces moteurs. Combien d'usagers sont capables de se représenter le mode de fonctionnement d'un moteur, d'en déduire que les causes de silence sont bien plus nombreuses que les sources de bruit et qu'une multitude d'informations pertinentes lui échappe ? La

maîtrise de ces outils est des plus aléatoires. S'ils offrent un service réel, ils ne peuvent prétendre organiser l'accès à des bases dispersées dès lors que celles-ci s'appuient sur des techniques de génération de pages dynamiques.

Bâtir un portail d'information documentaire

Le terme de portail d'information documentaire doit être précisé. Le portail désigne le point d'entrée unique d'un usager à un ensemble de ressources internes et externes. Il est associé à des fonctions d'identification, d'authentification et de gestion des droits d'accès en consultation, impression ou téléchargement. Pour devenir un portail d'information documentaire, il faut que l'outil intègre quelques fonctions supplémentaires : choix de bases à solliciter dans une liste fermée, choix des accès à utiliser, consultation simultanée des bases choisies, fusion des résultats de la recherche et si possible affichage de ces résultats par ordre de pertinence. Les portails d'information documentaire présentent l'intérêt d'accélérer considérablement la recherche : avec une seule requête, il est possible de solliciter un très grand nombre de bases. Si c'est un outil très efficace pour le professionnel de la documentation, plusieurs difficultés en limitent l'intérêt pour des usagers maîtrisant mal les techniques documentaires.

- D'abord, il faut déterminer les bases pertinentes par rapport à la recherche. Comme cela n'est pas facile, il est possible de solliciter toutes les bases mais le taux de bruit s'en ressent considérablement.
- Ensuite, il faut déterminer les accès performants par rapport aux bases sollicitées (mieux vaut-il rechercher par mot du titre ou par sujet ?). Cela suppose de connaître les traitements documentaires effectués dans ces bases (toutes effectuent-elles une indexation des sujets ? si ce n'est pas le cas et s'il faut se replier sur les mots du titre, dans quelles langues est-il utile de formuler la requête ? le français suffit-il ou mieux vaut-il tenter sa chance avec des termes anglais ?)
- Enfin, il faut savoir déterminer les termes de la requête et cela demande une connaissance encore plus fine du traitement documentaire (l'une des bases utilise-t-elle un répertoire encyclopédique comme Rameau et l'autre un thésaurus spécialisé ? faut-il employer des termes très généraux pour solliciter une base et des termes très précis pour l'autre base ?)

Le portail d'information documentaire devient acceptable pour des usagers peu familiarisés avec la documentation si les sites interrogeables sont présentés en des ensembles homogènes tant du point de vue des thèmes couverts que des traitements documentaires pratiqués. Ce n'est pas une mince affaire si le nombre de sites est très important. C'est là le paradoxe du portail : outil conçu pour faciliter l'accès à de très nombreux sites, il supporte mal l'utilisation par le public des primo-utilisateurs dès lors que se multiplient les sites proposés.

Tableau n°1. Quelques portails d'information documentaire

Fournisseur	Progiciels
Archimed	Book-Line
Dynix	Horizon Portail d'information avec RPA
Ever Team SA	ES Portal
Ex Libris	Metalib
Geac France	V-Link
Innovative Interfaces Inc.	MAP et GRE
Multilis Europe SA	iBistro/iLink, Rooms
Xerox askOnce	askOnce

Implanter dans l'opac un logiciel générant des liens contextuels

La toute dernière génération d'outils visant à mieux exploiter des ressources électroniques multiples et non coordonnées utilise la technique de génération de liens contextuels. Sans rentrer dans le détail de son principe de fonctionnement (notamment le standard OpenURL), la génération de liens contextuels s'appuie sur les postulats suivants :

- La plupart des usagers ne peuvent utiliser de manière efficace des sites multiples disposant de bases de références distinctes, aux logiques de consultation toujours différentes.
- La collecte d'informations sur le Web ne peut fonctionner de manière efficace que si la définition de l'objet recherché est très précise. Deux solutions peuvent être employées pour cette définition : mentionner un identifiant (ISBN, ISSN, ISRC, ISRN, ISMN, V-ISAN, ISTC, DOI...) ou mentionner des éléments bibliographiques précis (auteur+titre+éditeur...).

Intégrés au module de recherche documentaire ou à l'opac, les outils utilisant cette technique prolongent la consultation d'une référence par l'affichage de documents associés (résumé, analyse, critique, premier chapitre) ou par l'accès à divers services disponibles sur le Web. A l'issue de la consultation du catalogue, l'utilisateur n'a donc plus besoin de ressaisir les références de l'ouvrage ou de l'article l'intéressant pour accéder au texte du document ou à la commande de celui-ci en prêt entre bibliothèques ou directement auprès du libraire. Il n'a pas

non plus à choisir la base pertinente et peut raisonner simplement en termes de services ou d'informations recherchés (vais-je consulter plutôt le résumé ou lire le sommaire de cet ouvrage ?). En pratique on rencontre deux applications de cette technique : la collecte systématique d'informations et le rebond automatisé.

Première application : la collecte systématique d'informations apparentées à une référence

A l'issue de la consultation de l'opac, dès qu'une référence précise est affichée, le logiciel générant les liens contextuels collecte auprès de quelques fournisseurs prédéterminés des informations complémentaires. Il utilise pour cela les informations précises d'identification contenues dans la notice (et non les seuls termes de la requête de l'utilisateur). On peut imaginer ainsi la collecte du résumé et du sommaire dans la base du Cercle de la librairie pour un ouvrage français et dans une base type Books in Print pour des ouvrages anglais. L'utilisateur se voit présenter toutes ces informations comme si elles faisaient partie de la notice. Il n'a pas eu à demander une information précise, le logiciel affiche ce qu'il a pu trouver, c'est-à-dire peut-être le sommaire en l'absence d'un résumé.

Deuxième application : le rebond automatisé

L'utilisateur se tourne là aussi vers l'opac, saisit sa requête et se voit proposer, dès l'affichage d'une référence, un ensemble de fonctions lui permettant de prolonger sa recherche : un premier bouton donnera accès au premier chapitre de l'ouvrage, un deuxième au sommaire, un troisième à la table des matières... Cette application présente plusieurs avantages :

- l'utilisateur garde la maîtrise de la recherche d'information, il choisit entre le résumé et le sommaire ce qui lui paraît le plus intéressant,
- le logiciel, grâce à quelques règles de décision prédéfinies par la bibliothèque, va plutôt solliciter telle base Web s'il s'agit d'un ouvrage anglais, telle base Web s'il s'agit d'une thèse, telle autre base s'il s'agit d'un périodique.

Ces outils sont très prometteurs mais ils présentent tout de même un inconvénient majeur : ils reportent la complexité de la recherche sur la bibliothèque ou le centre de documentation qui aura dû prédéterminer les logiques de recherche que le logiciel mettra en œuvre. Pratiquement, cela repose sur la construction de règles de décision qui à partir du type de document, du secteur scientifique et d'éléments complémentaires définissent la liste des sites Web à solliciter pour obtenir un résumé, pour obtenir une table des matières... Si la bibliothèque arrive à déterminer des règles pertinentes et couvrant de manière satisfaisante l'ensemble du domaine traité, l'utilisateur n'aura plus à choisir les sites à interroger mais à retenir l'une ou l'autre des fonctions proposées « consulter le résumé », « consulter le premier chapitre »... Dans ces conditions, la génération de lien contextuel devient une technique extrêmement performante qui rend accessible une multitude de sites Web (ou de bases locales) à une très large population d'utilisateurs.

Tableau n°2. Quelques outils utilisant la génération de liens contextuels

Fournisseur	Progiciels	Application de collecte systématique d'informations	Application de rebond automatisé
Archimed	Book-Line avec Osc®		Oui
Dynix	Horizon Portail d'information	Oui	Oui
Ex Libris	SFX avec ou sans Metalib		Oui
Geac France	V-Link		Oui
Innovative Interfaces Inc.	MAP		Oui
Multilis Europe SA	iBistro/iLink	Oui	

Lien URL et génération de liens contextuels ont en commun l'enrichissement des fonctionnalités et des données proposées par l'opac. Moteur du Web et portail d'information documentaire sont des outils complémentaires de l'opac. Ces quatre solutions ne s'excluent pas, elles se complètent mais ne sont pas toutes adaptées à l'utilisation par une large population d'utilisateurs. Dans la réflexion sur son offre de services, la bibliothèque doit prendre en compte ces éléments. Une solution technique, ultra performante lorsqu'elle est mise en œuvre par des professionnels de la documentation, peut se révéler des plus décevantes lorsque l'on en confie l'utilisation à des utilisateurs inexpérimentés.

Tableau n°3. Le domaine d'emploi des quatre solutions facilitant l'accès du public à des ressources électroniques multiples et non coordonnées

	<i>Mieux articuler catalogue et sites Web en stockant l'adresse des sites en zone 856 de la notice Unimarc</i>	<i>Utiliser les moteurs de recherche du Web</i>	<i>Bâtir un portail d'information documentaire</i>	<i>Implanter dans l'opac un logiciel générant des liens contextuels</i>
Types de ressources se prêtant bien à l'utilisation de cette solution	Les encyclopédies. Des périodiques électroniques (en petit nombre). Tout autre document pour lequel il existe un site Web dont le contenu est le prolongement direct d'une publication imprimée.	Les sites Web gratuits composés de pages statiques (ce qui exclut le plus souvent les banques de données structurées, les catalogues...).	Des sites Web gratuits ou payants organisés en des ensembles homogène du point de vue des thèmes couverts, du traitement documentaire effectué et des langages documentaires utilisés.	Des sites Web gratuits ou payants conformes au standard Open URL (par exemple, Electre, la base du Cercle et de nombreuses bases de groupeurs).
Facteurs limitant l'efficacité de la solution	Les ressources électroniques peuvent être noyées dans la masse du catalogue qui contient bien souvent 99% de documents imprimés. L'établissement de ces liens est très lourd et parfois complexe.	Le bruit est tel qu'il fait oublier le silence. De plus, en l'absence de sélection préalable, l'utilisateur doit être capable de distinguer la valeur des différentes réponses (celle d'un organisme de normalisation est toute autre que celle contenue dans certaines pages personnelles).	L'utilisateur doit disposer d'une bonne représentation du contenu des bases, des traitements et des langages pour sélectionner les bases pertinentes et exprimer sa requête de manière efficace.	La bibliothèque doit avoir prédéfini des stratégies de recherche sous la forme d'un ensemble plus ou moins vaste de règles de gestion. La limitation tient à la lourdeur de l'expression de ces règles lorsque les sites Web deviennent très nombreux.