

# La recherche intelligente se redécouvre à l'Université Laval

## Un nouveau catalogue voit le jour



par Guy Teasdale  
Guy.Teasdale@bibl.ulaval.ca

L'auteur décrit les raisons ayant mené la bibliothèque de l'Université Laval à créer une nouvelle interface de recherche à facettes pour son catalogue, ainsi que les objectifs poursuivis. Il indique que l'outil aura des impacts positifs sur l'exploitation et l'utilisation du *Répertoire des vedettes matières de l'Université Laval* et de l'intelligence des métadonnées enfouies dans les catalogues. L'auteur fait sommairement état du débat concernant la pertinence des LCSH.

**Intelligent searching is rediscovered at l'Université Laval. A new catalogue sees the light of day** The author describes the rationale and objectives behind the Bibliothèque de l'Université Laval's decision to create a new faceted research interface for its catalogue. This tool will have a positive impact on the use of Université Laval's "*Répertoire des vedettes matières*" (RVM) and of the metadata buried in the catalogues. The author summarizes the debate relating to the pertinence of LCSH.

En 2005, 89 % des étudiants universitaires débutaient leur recherche d'information par un moteur de recherche de type Google ou Yahoo (contre 2 % dans un catalogue de bibliothèque)<sup>1</sup>. D'autre part, appelés à qualifier la recherche en bibliothèque par rapport à la recherche à l'aide d'un moteur de recherche, les usagers n'accordent la note de passage aux bibliothèques que pour deux des sept qualificatifs en cause : l'information est de qualité et elle est crédible. Les moteurs de recherche remportent la palme pour les cinq autres qualificatifs : rapide, pratique, facile à utiliser, rentable et fiable<sup>2</sup>.

Et cette vaste enquête a été publiée en 2005! Rien ne nous permet de penser que les opinions aient beaucoup évolué depuis ce temps. Les jeunes qui entrent à l'université sont pratiquement nés avec le Web, et les interfaces de recherche sont de plus en plus conviviales. Cette enquête de 2005 est corroborée par nos enquêtes LibQual qui font ressortir certaines insatisfactions des usagers au niveau du contrôle de l'information<sup>3</sup>.

Que faire alors? Attendre la prochaine version de notre système intégré de gestion de bibliothèques (SIGB)? Malheureusement, ce marché évolue beaucoup moins rapidement que toutes les applications commerciales du Web. Les SIGB génèrent des ventes d'environ 570 M\$ par année<sup>4</sup> tandis que Google à lui seul rapportait 5,3 G\$ au second trimestre 2008<sup>5</sup>! Les améliorations apportées dans les différents SIGB au fil des ans ont été qualifiées de « cosmétiques ». Un important auteur, Roy Tennant, parle même de « lipstick on a pig » (« rouge à lèvres sur un cochon », le « cochon » étant l'interface utilisateur). Certaines bibliothèques ont pris acte et tentent de fournir des interfaces mieux adaptées aux réalités du Web et de ses usagers.

C'est ainsi que la North Carolina State University Library (NCSU) provoqua un coup de tonnerre dans le ciel de la bibliothéconomie en dévoilant en janvier 2006 sa nouvelle interface de recherche propulsée par Endeca. Les blogueurs écrivent alors que la NCSU avait « tué le cochon » tant le saut qualitatif était important<sup>6</sup>. L'intérêt était double pour la bibliothèque de l'Université Laval (BUL), puisque ce nouveau catalogue de la NCSU permettait une recherche par facettes et se superposait au même SIGB, Unicorn, que celui de la BUL. Le saut qualitatif était important... et cher. Dans un texte précédent d'*Argus*, je mentionnais d'ailleurs que c'est sans doute là une des raisons qui poussent de plus en plus de bibliothèques à se « risquer » à l'Open Source<sup>7</sup>.

La Bibliothèque de l'Université Laval a donc examiné plusieurs options. Les choix ne manquaient pas : du côté des applications commerciales, on retrouve les Endeca, Primo

	Bibliothèque	Moteur de recherche
Rapide	8 %	92 %
Pratique	11 %	89 %
Facile à utiliser	15 %	85 %
Rentable	25 %	75 %
Fiable	28 %	72 %
Information de qualité	56 %	44 %
Information crédible	60 %	40 %

(Ex Libris), Encore (Innovative), Aquabrowser, sans compter les autres « en développement ». Du côté de l'Open Source, on retrouve Evergreen, Fac-Back, Scriblio, VuFind, Koha, PMB et BiblioteQ. Après analyse, le logiciel SOLR nous a semblé être le plus prometteur. À l'automne 2007, un prototype contenant 1,4 million de notices (extraites d'Unicorn et converties en XML) a été développé, et différents tests ont été effectués. Le projet, présenté aux employés, a alors été jugé emballant. Disposant de toutes les expertises professionnelles nécessaires à l'interne<sup>8</sup>, la BUL a ainsi décidé d'aller de l'avant en mettant sur pied un projet de développement en bonne et due forme. Et c'est ainsi que « Ariane 2.0 » a été dévoilé dans sa version bêta le 13 août dernier (<http://ariane2.bibl.ulaval.ca>).

## Objectifs

Les objectifs poursuivis étaient nombreux. D'abord et avant tout, il va sans dire, mieux répondre aux besoins des usagers et régler les problèmes signalés au début de cet article. Dès le départ, nous avons choisi de maintenir les deux catalogues Ariane, ce qui diminuait les possibilités de résistance au changement.

Nous cherchions à recréer le « rectangle magique » dépouillé de Google et à l'appliquer à une bibliothèque universitaire. Les avantages sont tangibles, les plus évidents étant la recherche par facettes; la possibilité de mieux exploiter la puissance du *répertoire des vedettes matières de l'Université Laval* (RVM) et des différents fichiers d'autorité; la possibilité de trier des résultats de recherche de différentes façons; la pondération des résultats de recherche en fonction de la présence de mots dans telle ou telle zone du format MARC;

« Ariane 2.0 tiendra compte de la popularité des documents lors de l'affichage des résultats. »

la pondération des notices en fonction de la « popularité » du livre (nombre de prêts); la prise en compte des mots vides dans plusieurs langues; la possibilité d'indexer la racine des mots; la possibilité d'intégrer un correcteur orthographique... et il y en a d'autres!

Des sondages et des *focus groups*, visant à mesurer la facilité d'utilisation et la satisfaction par rapport à l'ergonomie, seront effectués au cours de l'automne 2008 et permettront l'évolution rapide de la nouvelle plateforme de recherche.

## Guider la recherche avec les facettes

Nous cherchons à guider l'étudiant dans sa démarche de recherche en facilitant une recherche booléenne intuitive. Car il faut se rendre à l'évidence : cela fait près de trente ans qu'on explique les principes de base de la recherche booléenne à des générations d'étudiants et la littérature bibliothéconomique démontre à l'envie que cela pose toujours

problème. Soit nos étudiants ne comprennent pas, soit ils ne voient pas l'intérêt de la chose. À l'opposé, les facettes d'Ariane 2.0, construites dynamiquement au fur et à mesure d'une recherche, permettent de raffiner les résultats d'un simple clic. Il est également facile d'étendre une recherche, en revenant dans des étapes antérieures de la recherche. En effet, les différentes facettes activées lors d'une même recherche peuvent être annulées, dans n'importe quel ordre.

## Passer du centre du campus au centre de l'écran

Nous voulons également faciliter l'intégration harmonieuse du catalogue aux futurs environnements numériques d'apprentissage en train d'être implantés sur les campus, afin que chaque étudiant soit à un clic de la bibliothèque depuis son environnement d'apprentissage.

Dans une étape ultérieure, nous évaluerons l'intérêt de permettre l'interaction avec l'utilisateur (évaluations d'utilisateurs, comptes rendus, *tagging*, c'est-à-dire mots clés d'utilisateurs). Comme la BUL contrôle maintenant son interface, elle pourra implanter ces améliorations à son rythme, en fonction des besoins d'utilisateurs, et non à la suite de développements éventuels et coûteux de la part des fournisseurs commerciaux.

## Redécouvrir la puissance du RVM et des autorités

Plusieurs débats, parfois virulents<sup>9</sup>, agitent nos collègues américains sur la pertinence de maintenir les fameuses Library of Congress Subject Headings (LCSH). Les partisans de leur abandon font valoir les arguments des coûts, d'une utilisation déclinante et d'un usage de plus en plus important des Google de ce monde.

Étant producteur du RVM, la BUL est, bien évidemment, interpellée par la nature du débat. Ariane 2.0 a été conçue avec l'intention d'en faire un outil permettant une utilisation plus efficace et harmonieuse des fichiers d'autorité et du RVM. Nous espérons faire ressortir la richesse et l'intelligence des métadonnées enfouies dans le catalogue par des générations de bibliothécaires. Ainsi, une recherche d'un terme rejeté ou même d'un terme anglais pourra offrir différentes options à l'utilisateur en plus des termes génériques, spécifiques et associés.

Exemples : dans l'index sujets/répertoire

Vous avez demandé : Guerre fratricide

La bibliothèque de l'Université Laval vous suggère :

Terme générique : Droit international

Terme générique : Guerre

Terme générique : Résistance au gouvernement

Terme générique : Révolutions

Forme acceptée : Guerre civile

Exemple : dans l'index auteur/répertoire

Vous avez demandé : Amantine Dupin

La bibliothèque de l'Université Laval vous suggère :

Sand, George, 1804-1876

Exemple : dans l'index auteur/Titre normalisé de l'oeuvre  
Vous avez demandé : Mozart petite musique de nuit  
La bibliothèque de l'Université Laval vous suggère :  
Mozart, Wolfgang Amadeus, 1756-1791. Kleine Nachtmusik

### Qu'a-t-on à perdre?

L'exploitation du RVM dans Ariane 2.0 permettra à l'étudiant de découvrir, au fil de ses recherches, deux importantes valeurs ajoutées par rapport à une simple recherche par mot-clé, soit la prévisibilité et ce que Thomas Mann appelle la « sérendipité systématisée »<sup>10</sup>.

*« On redécouvre les vedettes matières  
précoordonnées ! »*

En cherchant un mot dans l'index « sujet/répertoire », on accédera aux vedettes matières précoordonnées comme on les voyait dans les catalogues sur fiches. L'étudiant y découvrira, un peu intuitivement, une structure intellectuelle avec différentes subdivisions par ordre alphabétique. Cette présentation fera surgir des aspects de recherche auxquels l'utilisateur n'avait sans doute pas pensé. Il découvrira progressivement, à l'usage, que cette structure est prévisible. Par exemple, une recherche sujet de l'expression « guerre civile » suggérera un arrangement par ordre alphabétique de pays avec subdivisions de genres ou de formes comme des biographies, des sources et des récits personnels.

La sérendipité (le RVM nous renvoie à « effet serendip ») se définit comme une façon systématisée de faire des « découvertes heureuses ou inattendues ». Le meilleur exemple réside dans le bouquinage dans les rayons de bibliothèque, permettant de trouver des livres intéressants à proximité de celui que l'on cherchait. Elle devient systématisée en ce sens que le livre n'est pas placé au hasard, mais plutôt par sa cote. De la même façon, une recherche sur la « peine capitale » dans la zone titre d'Ariane 2.0 générera 14 titres. La même recherche dans la zone sujet/répertoire permettra de voir que la vedette matière acceptée est plutôt « peine de mort ». La relance de la recherche avec ce concept fera ressortir pas moins de 152 titres, permettant des découvertes heureuses ou inattendues.

Ces deux avantages me semblent démontrer de façon éloquente la pertinence du RVM. Une étude récente démontre qu'en utilisant uniquement une recherche par mot clé on passerait à côté d'environ 30 % des ouvrages<sup>11</sup>! On perdrait également les avantages de l'entrée uniforme peu importe la langue (colocalisation). Les Américains ne repéreraient que les ouvrages en anglais et inversement les francophones interrogeant en français passeraient à côté d'une bonne partie de la collection en d'autres langues, ce qui, à la BUL, représente tout de même près de 50 % de la collection!

### Et si on « cherchait partout », comme avec Google?

Certains partisans de l'abandon des LCSH prétendent que les recherches en catalogue sont surtout faites pour des choses connues d'avance. Dans un exemple cité par l'Université Cornell<sup>12</sup>, on utilise les champs auteur et titre dans 61 % des cas, contre 3 % seulement pour les recherches sujet. Notons qu'à l'Université Laval, c'est plutôt 11 % des recherches qui sont faites par sujet, ce qui illustre davantage encore la nécessité de la colocalisation pour des usagers francophones.

D'autre part, on constate que les étudiants utilisent Ariane/Unicorn de la même façon que Google : ils saisissent un ou deux mots clés en vocabulaire libre et recherchent dans tous les champs. Après avoir trouvé une notice intéressante, ils relancent la recherche à partir d'hyperliens présents dans les champs en vocabulaire contrôlé de la notice bibliographique. Ainsi, à partir des 3,9 M de recherches effectuées dans Ariane l'année dernière, quelque 250 000 relances se sont produites à partir d'une notice de catalogue et 527 000 autres, à partir d'utilisation de termes présents dans les fichiers d'autorité. Près d'une recherche sur cinq (19,8 %) est ainsi effectuée par l'entremise du vocabulaire contrôlé.

Étonnamment, après un an et demi d'utilisation, la NCSU a calculé que les recherches au moyen de la facette sujet représentent 28,5 % des requêtes à son catalogue. Il s'agit là de la facette la plus utilisée, suivie de la classification de la bibliothèque du Congrès (24,3 %)<sup>13</sup>. Nous croyons que l'impact sera le même pour Ariane 2.0.

Depuis la tablette d'argile, les besoins changent, les bibliothèques s'adaptent et les outils se raffinent. Ariane 2.0 participe à ce mouvement sans fin. D'autres défis s'annoncent : par exemple, ceux des coûts du traitement intellectuel consécutif à l'explosion documentaire, la réutilisation des différentes formes de métadonnées, l'expérimentation du 2.0 à travers des outils de type LibraryThing, lequel enregistre pas moins de 460 000 usagers ayant catalogué 30 M de livres au moyen de 38 M de mots-clés!

À l'heure où l'on s'apprête à construire le Web sémantique annoncé en 2001 par Tim Berner-Lee, à l'heure où l'on cherche désespérément des sources crédibles de vocabulaire contrôlé à exploiter par machine (et que les bibliothèques possèdent déjà ce capital de qualité et de crédibilité), il serait catastrophique, pour ne pas dire stupide, d'abandonner les LCSH, ce grandiose édifice intellectuel construit depuis plus de cent ans par des générations de bibliothécaires. Ariane 2.0, nous en sommes convaincus, fera redécouvrir l'art de la recherche intelligente.

### Notes

1. Perceptions of libraries and information resources, OCLC, 2005, question 520 ([http://www.oclc.org/reports/2005\\_perceptions.htm](http://www.oclc.org/reports/2005_perceptions.htm)).
2. Source: Perceptions of libraries and information resources, OCLC, 2005, question 1355. (<http://www.oclc.org/reports/2005perceptions.htm>)

3. Source : LibQUAL+ Canada Consortium (rev. 9/17/07). Results from all 54 participating libraries (<http://library.queensu.ca/webir/canlibqual/results-e.htm>). Question Information Control # 6 «Easy to use access tools that allow me to find things on my own». Les réponses étaient : moyenne minimum : 6.88; moyenne désirée 8.10; moyenne perçue 7.04.
4. Breeding, Marshall. «Opportunity out of turmoil». *Library Journal*, vol. 133, no. 6 (April 2008): 32-38.
5. «Google announces second quarter 2008 results». *Business Wire*, July 17, 2008 ([http://findarticles.com/p/articles/mi\\_m0EIN/is\\_2008\\_July\\_17/ai\\_n27926941](http://findarticles.com/p/articles/mi_m0EIN/is_2008_July_17/ai_n27926941)).
6. Koltzenburg, Teresa. «From swine to divine: NCSU unveils new online catalog», 12/01/2006. (<http://www.techsource.ala.org/blog/2006/01/from-swine-to-divine-ncsu-unveils-new-online-catalog.html>).
7. Teasdale, Guy. «Les vannes de l'Open Source sont-elles ouvertes?». *Argus*, vol. 36, no2, automne 2007.
8. Simon Pierre Barrette, spécialiste en ressources documentaire, Jo-Anne Bélair, chef de la section Répertoire des vedettes matières, Rida Benjelloun, chef de la section recherche et développement numérique et chef de projet, Pierre Lemieux, analyste de l'informatique, Marcel Plourde, chef de la section catalogage, Chantal St-Louis, chef de la section Aide à la recherche et Guy Teasdale, directeur de la bibliothèque numérique et des TI. Ces personnes étaient assistées d'un comité d'utilisateurs formé d'Élise Couturier, Annie D'Anjou, Louise Delisle et France Morin.
9. Voir en particulier les interventions de Thomas Mann, bibliothécaire à la Bibliothèque du Congrès. Mann, Thomas. «What is going on at the Library of Congress?» (<http://www.guild2910.org/AFSCMEWhatIsGoingOn.pdf>). Mann, Thomas. «The changing nature of the catalog and its integration with other discovery tools. Final Report. March 17, 2006. Prepared for the Library of Congress by Karen Calhoun : A Critical Review» (<http://www.guild2910.org/AFSCMECalhounReviewREV.pdf>).
10. Mann, Thomas. *Library Research Models : A Guide to Classification, Cataloging, and Computers*. New York, Oxford University Press, 1993, p.120.
11. Gross, Tina et Arlene G. Taylor. «What have we got to lose? The effect of controlled vocabulary on keyword searching results». *College & Research Libraries*, May 2005, pp. 212-230.
12. Banush, David et Jim LeBlanc «Utility, library priorities, and cataloging policies». *Library Collections, Acquisitions & Technical Services*, 31, 2007, p. 106
13. «Which facets are most used?». Courriel d'Emily Lynema à la liste de discussion NGC4LIB@listserv.nd.edu, 8 avril 2008.

Directeur, Bibliothèque numérique et technologies de l'information, Bibliothèque de l'Université Laval

## Programme de formation continue 2008/2009 de la CBPQ

- **17 octobre** – ENAP *Mettre en place des pratiques de transmission des connaissances dans son milieu de travail*
- **31 octobre** – CRIM *Utilisation et production de fils RSS en bibliothèque* – Atelier technique
- **7 novembre** – ENAP *La littérature pour les 12-17 ans, un monde à découvrir* – Atelier d'initiation
- **14 novembre** – CRIM *Design, développement et production de capsules vidéo didactiques* – Atelier technique
- **28 novembre** – CRIM *Intégrer les blogues à sa pratique : atelier de création et d'exploitation* – Atelier technique
- **6 février 2009** – *Planifier ses stratégies pour mieux se positionner !*
- **13 février** – *Intégrer les wikis à sa pratique : atelier de création et d'exploitation* – Atelier technique
- **20 février** – *Le Web, les jeunes et la bibliothèque*
- **13 mars** – *Comprendre les liens entre la politique municipale, l'administration et la bibliothèque publique*
- **20 mars** – *La gestion des employés difficiles*
- **27 mars** – *Le Mind mapping® ou les cartes cognitives au service des bibliothèques et centres de documentation*
- **3 avril** – *Les collections dans les bibliothèques universitaires : élaborer et mettre à jour une politique de développement des collections*

Pour plus d'informations :

[info@cbpq.qc.ca](mailto:info@cbpq.qc.ca)

[www.cbpq.qc.ca/formation\\_continue/fc\\_2008-2009/FC0809\\_Programme.html](http://www.cbpq.qc.ca/formation_continue/fc_2008-2009/FC0809_Programme.html)