Étude pour la mise en place d'un moteur de recherche globale pour le " knowledge center " de SKEMA Business School

Christophe Willaert

▶ **To cite this version:**

Christophe Willaert. Étude pour la mise en place d'un moteur de recherche globale pour le " knowledge center " de SKEMA Business School. domain_shs.info.docu. 2010. mem_00502620

HAL Id: mem_00502620

https://memsic.ccsd.cnrs.fr/mem_00502620

Submitted on 15 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Christophe Willaert
Master 1, mention ICD
(Option : Sciences de l'Information et du Document)

MÉMOIRE DE STAGE
Mission effectuée du 12 avril au 11 juin 2010
à SKEMA Business School, Lille

Étude pour la mise en place d'un moteur de recherche globale pour le « knowledge center » de SKEMA Business School

Sous la direction de :

Sophie Chauvin (maître de conférence en SIC, responsable universitaire)
Isabelle Normand (community manager TICe , responsable professionnelle)

Année universitaire 2009/2010

REMERCIEMENTS

Je tiens tout d'abord à remercier Mme Isabelle Normand, responsable de la médiathèque « la Fusée » au sein de SKEMA Business School, de m'avoir accueilli au sein de son équipe. Je tiens aussi à la remercier pour le temps qu'elle a eu l'amabilité de me consacrer.

Au sein de l'université Lille 3, je souhaite adresser mes sincères remerciements à Mme Sophie Chauvin, maître de conférence en SIC, pour son aide, sa disponibilité, mais aussi et surtout pour la confiance dont elle a fait preuve à mon égard.

Merci à Déborah Potelle, Maria Da Silva, Aurélie Duclos et France Lefebvre du Prey, pour leur accueil, leur soutien et leur grande disponibilité.

Je voudrais décerner une mention spéciale à Karine Boyer pour son dévouement quant à la relecture de mon mémoire et qui a été la première à me donner son avis, ainsi qu'à Yann Griset et Charlotte Varin pour leurs conseils pertinents et leur bonne humeur.

Je remercie également l'ensemble des membres de SKEMA avec lesquels j'ai eu l'occasion d'échanger, en particulier les personnes du service informatique qui ont pu me dépanner plus d'une fois.

Je tiens enfin à remercier M. Joachim Schopfel, directeur de l'UFR IDIST qui nous a fait le plaisir de venir participer à la journée d'inauguration de « la Fusée ».

TABLE DES MATIERES

REMERCIEMENTS **2**

TABLE DES MATIERES **3**

TABLE DES FIGURES **4**

1 INTRODUCTION **6**

2 CONTEXTE **8**

2.1 SKEMA BUSINESS SCHOOL 8
2.2 LE CAMPUS DE LILLE 9
2.3 LA FUSEE 10
2.3.1 Présentation générale 10
2.3.2 Public et missions 12
2.3.3 L'accès aux ressources 13
2.4 LE PROJET DE KNOWLEDGE CENTER 19
2.4.1 Étude des besoins 19
2.4.2 Les ressources documentaires 21

3 LA RECHERCHE D'INFORMATION EN LIGNE 22

3.1 LES STRATEGIES DE RECHERCHE DOCUMENTAIRE DES ETUDIANTS 22
3.2 LES SYSTEMES DE RECHERCHE D'INFORMATION 25
3.3 ASPECTS COGNITIFS DE LA RECHERCHE D'INFORMATION 28
3.4 VISUALISATION ET INTERFACES DE RECHERCHE 29
3.4.1 Le rôle de l'ergonomie pour les bibliothèques numériques 29
3.4.2 Exemples d'interfaces pour l'accès et la navigation dans les catalogues en ligne 31
3.4.3 Le recours aux langages de pictogrammes et aux schémas 44
3.5 L'UTILISATION DE FACETTES POUR L'ACCES AUX RESSOURCES 52

4 LES BIBLIOTHEQUES A L'HEURE DU NUMERIQUE 57

4.1 VERS UNE BIBLIOTHEQUE « 2.0 » 57
4.2 LES PORTAILS DOCUMENTAIRES 58
4.3 LES OPACS DE NOUVELLE GENERATION 60

5 LES SOLUTIONS DE RECHERCHE GLOBALE 63

5.1 LA RECHERCHE FEDEREE 63
5.2 LA RECHERCHE UNIFIEE 66

6 CONCLUSION ET PERSPECTIVES 71

7 REFERENCES 74

TABLE DES FIGURES

Figure 1 : Interface pour requêter le moteur AquaBrowser 13
Figure 2 : Interface de résultats du moteur AquaBrowser 14
Figure 3 : Menu d'accès aux ressources numériques 15
Figure 4 : Fiche de présentation du service Business Digest 15
Figure 5 : Page d'accès aux bases de données 16
Figure 6 : Recherche orientée sous forme de carte arborescence 16
Figure 7 : Liens pour l'accès direct au site Factiva 17
Figure 8 : Système actuel de recherche (multiples points d'entrée) 18
Figure 9 : Système de recherche unifiée (un seul point d'entrée) 20
Figure 10 : Différences significatives des méthodes de découverte d'articles en fonction de l'âge 23
Figure 11 : Les quatre grands types de besoin en information (d'après Morville, 2007) 26
Figure 12 : Les différents modes et types de recherche, d'après Gilles Balmisse (2008) 27
Figure 13 : L'anatomie de base d'un système de recherche, d'après Morville (2007) 27
Figure 14 : Modèle des attributs de l'acceptabilité d'un système 31
Figure 15 - La fonctionnalité "Roue magique" de Google 33
Figure 16 : La recherche d'images avec Google Image Swirl 34
Figure 17 : Interface de navigation de terme à terme, avec conservation des traces de navigation 34
Figure 18 : Recherche par constellations sur le site de la bibliothèque municipale de Lille 35
Figure 19 : L'interface du moteur EyePlorer 36
Figure 20 : Découpage circulaire de l'information avec Crystal Mapping 36
Figure 21 : Carrot2 et ses clusters sous forme d'arbre hyberbolique 37
Figure 22 : Wikipedia Roll 37
Figure 23 : Une visualisation en cluster au sein du portail KmP 37
Figure 24 : Interface du moteur KartOOvisu 38
Figure 25 : Grokker : technique de représentation par pavage de cercles 39
Figure 26 : Navigation à facettes dans les articles du New-York Times avec Elastic Lists 39
Figure 27 : Système de navigation dans les articles du projet MACE 40
Figure 28 : L'interface de recherche visuelle d'EBSCO Host 41
Figure 29 : Evri, graphe des connections et diagramme des tendances 41
Figure 30 : Deux fonctionnalités de SiloBreaker : filtrage par facettes avec icônes et graphe de relations 42
Figure 31: Indicateurs visuels, placés dans la colonne de droite, sur le moteur des Echos 42
Figure 32 : Constellations, projet d'Exalead 43
Figure 33 : Le projet Virtual Shelf Index de la bibliothèque de l'Université d'État Caroline du Nord 44
Figure 34 : UpToTen et son interface iconique 46
Figure 35 : Symbaloo 46
Figure 36 : Plan d'une fusée, tirée de la bande-dessinée « On a marché sur la Lune » (Hergé) 47
Figure 37 : Exemple de représentation d'un bâtiment selon une coupe horizontale 47
Figure 38 : Greater Shakespeare Map, par la Royal Shakespeare Company 47
Figure 39 : ConceptSpace, un thésaurus visuel pour trouver des documents sur des sujets précis 48
Figure 40 : ConceptSpace, page de sélection des bases sur un terme donné 48
Figure 41 : L'interface du Visual Catalog tel que présenté sur le site de l'Université d'Artois 49
Figure 42 : VQuery (Jones and McInnes, 1998) interface à base de diagrammes de Venn pour la spécification de requêtes booléennes. 50
Figure 43 : Issuu et son interface de consultation de magazines 51
Figure 44 : Le nouveau lecteur PDF intégré à la plateforme EBSCOHost 52
Figure 45 : Interface de lecture de Google Books 52
Figure 46 : Exemples de facettes proposées par différentes bibliothèques (Maisonneuve et Touitou, 2007) 53
Figure 47 : Page de résultats d'AquaBrowser sur le portail de la médiathèque de Skema BS / ESC Lille 55
Figure 48 - AquaBrowser : Notice d'un ouvrage 56
Figure 49 : L'articulation des principales briques d'un portail pour bibliothèque (d'après Maisonneuve, 2007) 59
Figure 50 : L'utilisation d'un OPAC de nouvelle génération (Maisonneuve, 2009) 61
Figure 51 : L'architecture d'un OPAC de nouvelle génération (Maisonneuve, 2009) 62

Figure 52 : Schéma représentant le principe de la recherche fédérée _____ *63*
Figure 53 : Architecture fonctionnelle de services OAI pour la recherche unifiée _____ *67*
Figure 54 : Statistiques d'usage des bases à Grand Valley avant et après le passage à Summon _____ *69*
*Figure 55 : Page de résultats de Summon*_____ *70*

1 INTRODUCTION

Première école de commerce de France en nombre d'étudiants, SKEMA Business School est née de la fusion du groupe ESC Lille et du CERAM[1] de Nice, en novembre 2009. Avec pour objectif de former les managers de demain, elle fait face aux défis de l'innovation permanente. Transformé par les nouvelles modalités d'apprentissage et les potentialités pédagogiques des outils numériques, le campus de Lille a commencé très tôt à penser et mettre en place une bibliothèque 2.0, un écosystème pour co-apprendre, travailler et innover. Aujourd'hui, la volonté de SKEMA Business School est de poursuivre dans cette voie avec la mise en place d'un « knowledge center » qui doit permettre à terme d'accéder à la globalité des connaissances issues des différents campus : le fonds documentaire des bibliothèques, les contenus des outils d'elearning, ainsi que les informations produites par les étudiants (blogs, wikis, sites, …). Le moteur de recherche devient ainsi le point clé d'accès aux données et à l'information.

Plusieurs études (Assadi et Beaudouin, 2002, Griffiths et Brophy 2005, Martin et Perret, 2009) ont montré que les étudiants n'utilisent en général qu'un seul moteur de recherche sur le Web et qu'ils ont tendance à transposer les habitudes et les routines acquises au catalogue en ligne qui leur est proposé. Il m'a été demandé d'étudier quelles solutions peuvent être envisagées pour répondre à ces attentes, avec une approche qui mise sur la simplicité, qui permet un accès via différentes interfaces (intégrée au portail, sous forme de « widget », pour terminaux mobiles, ...) et qui s'intégre aux modes et contrôles d'accès existants. Les fonctionnalités sociales qui permettent le partage et la recommandation, voire l'ajout de commentaires, doivent également être prises en compte. On peut résumer les besoins des utilisateurs comme suit :

- Retrouver l'information pertinente à travers des interfaces qui sont simples, intuitives et homogènes ;
- Retrouver des documents répondant à des critères spécifiques tels qu'une date, une langue, un type, etc. ;

[1] Centre d'Enseignement et de Recherche Appliqués au Management <http://fr.wikipedia.org/wiki/CERAM_Sophia_Antipolis_-_ESC_Nice>

- Avoir des résultats de recherches présentés dans un format simple et facile à comprendre ;
- Visionner et télécharger une partie de tous les documents retrouvés.

Afin de bien distinguer les outils de gestion interne, de recherche unifiée et les interfaces de service, j'ai effectué une analyse fonctionnelle de chaque besoin. L'étude des sources d'information à intégrer a montré que nous sommes face à deux catégories de données à indexer : celles issues de fonds documentaire (imprimé et numérique), relevant du domaine de la bibliothéconomie (tels que les services de recherche fédérée), et les autres informations, issus des outils d'apprentissage et de type web 2.0, relevant plus des moteurs de recherche traditionnels que l'on trouve en entreprise. C'est la partie « bibliothéconomique » qui va être traitée dans ce mémoire, celle-ci étant la première étape abordée pour l'étude d'une solution.

2 CONTEXTE

2.1 SKEMA Business School

SKEMA est l'acronyme de School of Knowledge Economy and MAnagement. L'objectif de l'école avec ce nouveau nom est de confirmer sa lisibilité à l'international. SKEMA Business School compte aujourd'hui plus de 6000 étudiants, répartis sur cinq sites :

- en France (Lille, Paris La Villette, Paris La Défense, Sophia Antipolis)
- en Chine (Suzhou)
- au Maroc (Casablanca)

Viendront prochainement s'ajouter des campus aux États-Unis (à Raleigh, en Caroline du Nord , pour la rentrée 2010), en Inde et au Brésil. SKEMA totalise :

- 6245 étudiants (1ère école de commerce française) ;
- 138 professeurs permanents (2ème école de commerce française) ;
- 55 millions d'euros de budget (3ème école de commerce française) ;
- 20000 diplômés dans le monde entier ;
- 268 publications, 70 étoiles CNRS ;
- Plus de 65 associations d'étudiants

Les objectifs statutaires de l'association loi 1901 SKEMA sont centrés sur les étudiants et concernent l'enseignement, la recherche en management ainsi que l'ouverture internationale. Ils intègrent aussi la participation au développement économique, social, scientifique et culturel des territoires.

L'école est certifiée ISO AFAQ, norme qui définit des standards de management autour de grands principes comme l'amélioration continue, l'orientation client, le management par les processus… C'est aussi la première école de commerce non américaine à avoir reçu l'accréditation internationale GAC PM (Global Accreditation Center for Project Management) en 2005.

2.2 Le campus de Lille

L'ESC Lille a été fondée en 1892. Le campus, désormais situé à deux pas de la gare Lille Flandres, dans le quartier d'affaires de la métropole, offre des formations aux niveaux bachelor, master et doctorat. La concurrence dans la région entre les établissements de formation en management est forte, avec la présence notamment de l'EDHEC, de l'IESEG et de l'IAE de l'Université de Lille 1. En dix ans, les effectifs étudiants sont passés de 1500 en 1999 à près de 3000 en 2009.

Après avoir constaté que la moitié des étudiants vient de la région parisienne, l'ESC Lille a ouvert un autre site à Paris, dans le quartier de la Villette. Outre ses liens privilégiés avec des écoles d'ingénieurs (joint-venture avec Centrale Lille) ou des universités françaises (Lille 2, Littoral...), le groupe a développé 66 partenariats internationaux.
Trois programmes principaux sont offerts par l'école :

- le programme ESDHEM (environ 15% des effectifs). Il constitue une licence et un programme de préparation aux concours grandes écoles.
- le programme Grande Ecole [PGE] (environ 60% des effectifs). Le titre est visé par l'Etat et octroie le grade de Master.
- le programme Post-Graduate (environ 30 % des effectifs). Diplômes de l'école : Mastères spécialisés, M.B.A., Ph.D.

Ces dernières années, l'établissement s'est doté de nouveaux outils pour favoriser le travail en mode projet. C'est ainsi qu'a été mis en place un ENT[2] basé sur la solution *Google Apps Education*[3], complété par une plateforme d'e-learning (*Claroline*[4] et *CrossKnowledge*[5]).

En août 2009, l'AERES (Agence d'Évaluation de la Recherche et de l'Enseignement Supérieur) a publié un rapport[6] sur l'École supérieure de commerce de Lille ; évaluation effectuée avant la fusion avec le CERAM. Ce rapport parle d' *« une organisation ouverte et marquée par la forte implication du monde socio-économique, centrée sur des objectifs de*

[2] Environnement Numérique de Travail
[3] http://www.google.com/a/help/intl/fr/edu/index.html
[4] http://www.claroline.net/
[5] http://www.crossknowledge.net/
[6] Rapport de l'établissement ESC Lille - Campagne AERES 2008-2009 (Vague D 2010-2013) <http://www.aeres-evaluation.fr/Etablissements/ESC-Lille>

professionnalisation et d'employabilité et où les objets principaux sont bien les objets pédagogiques, en particulier le PGE » et précise que *« les étudiants rencontrés se disent très fiers d'appartenir à cette école et très satisfaits de ce qu'ils y acquièrent »*.

2.3 La Fusée

2.3.1 Présentation générale

La médiathèque du campus de Lille est gérée par six « Geemiks »[7] qui sont médiathécaires et animatrices de communauté. Isabelle Normand, responsable de la médiathèque depuis 1994 exerce également depuis février 2009 la fonction de *community manager TICe*[8] qui l'implique davantage dans le déploiement de la stratégie TICe, avec la responsabilité du système d'information documentaire de l'école. Ainsi en mai 2009, les Geemiks ont lancé le projet *Youontheweb*[9] pour enseigner les bonnes pratiques en gestion d'identité numérique auprès des étudiants.

Bénéficiant de l'agrandissement récent de la surface du campus de Lille, la médiathèque a migré en novembre 2009 dans de nouveaux locaux de 1000 m^2 (trois fois plus grand que le précédent lieu). Baptisé « la Fusée », ce nouvel espace a été pensé pour *« favoriser la connectivité humaine et technologique afin de multiplier les échanges d'informations, augmenter les contributions, favoriser le partage des expériences, des intérêts, des passions, des rêves et des projets entre toutes les personnes »*[10]. La Fusée correspond au modèle de

[7] *« Geemik est un néologisme composé de geek signifiant férue de nouvelles technologies et gimmick un terme de jazz désignant les cinq premières notes qui donne le ton et capte l'oreille de l'auditeur. Le geemik propose des infos, outils, des pistes de réflexion utiles pour la communauté. Il teste les outils, forme les membres à de nouveaux usages, imagine des développements et des applications intéressantes. Le geemik donne le gimmick à sa communauté mais laisse les participants les interpréter à leur rythme. Les membres du réseau commentent, évaluent, s'approprient ou non les différentes ressources. Le geemik n'est pas le chef d'orchestre. Il est un des musiciens du jazz band. Les contributions et les échanges entre tous les joueurs enrichissent la partition initiale. »* Source : Dico du futur
http://www.dicodufutur.com/Geemik_a168.html

[8] Technologies de l'Information et de la Communication pour l'éducation. Ce terme recouvre les outils et produits numériques pouvant être utilisés dans le cadre de l'éducation et de l'enseignement.

[9] http://youontheweb.fr/

[10] *Du rêve à la réalité : naissance d'un nouvel espace connecté La Fusée*, blog des Geemiks, 19 novembre 2009
http://www.geemik.net/?p=450

« carrefour de l'information »[11] (*information commons* ou *learning commons*), issu d'une réflexion menée à la fin des années 1990 par des gestionnaires de collèges et d'universités américaines se questionnant sur l'avenir de la bibliothèque de leur établissement. Ce terme recouvre un espace physique, généralement situé dans les bibliothèques, avec :

- des postes de travail informatiques équipés de logiciels pouvant supporter une variété d'utilisations ;
- des espaces réservés au travail individuel ou de groupe ;
- du mobilier confortable ;
- du personnel apte à soutenir les activités reliées à l'accès à l'information et à l'utilisation des technologies.

L'intégration entre l'enseignement, l'acquisition de connaissances, la documentation et la formation aux technologies est au cœur de cette notion qui renouvelle la conception de la relation entre formation et bibliothèques. Elle réduit les frontières entre enseignement et documentation et permet des modes de travail dynamiques et partagés (travail de groupe et production de documents, souvent multimédia, étant vivement encouragés) (Jouguelet, 2009). Sur le plan architectural, une des caractéristiques des *learning centres* est la différenciation de zones selon les modes de travail : travail individuel au calme, zone de silence, ou travail de groupe (ibid).

Repenser la bibliothèque en favorisant l'émergence d'usager-acteur, tel est le crédo des Geemiks. Le portail actuel de la médiathèque[12], lancé en décembre 2007, illustre ces changements vers une « bibliothèque 2.0 ». Il intègre le moteur de recherche visuel à facettes *AquaBrowser*[13], un système de chat (*Meebo*[14]) pour permettre aux étudiants d'interagir avec les Geemiks, ainsi que des univers thématiques *Netvibes*[15] dédiés à chaque communauté :

[11] Catherine Lamy, *au Québec, la bibliothèque, un lieu pour demain*, Archimag – Guide pratique « la bibliothèque à l'heure du web 2.0 », juillet 2008
Dans son rapport sur les *learning centres*, Jouguelet (2009) propose la traduction de « centre de ressources pour l'information et la recherche ».
[12] http://mediatheque.lafusee.fr/
[13] http://aqua.esc-lille.eu/
[14] http://www.meebo.com/
[15] http://fr.wikipedia.org/wiki/Netvibes

- Droit, fiscalité et économie
- Finance, contrôle et audit
- Logistique et achats
- Management de projets et programmes
- Marketing, e-commerce et communication
- Management et ressources humaines
- Stratégie et management entrepreneurial

Pour faciliter l'accès aux ressources, les Geemiks travaillent avec des équipes étendues (informatique, pédagogie, communication, …) et mettent en place des produits à valeur ajoutée pour leurs communautés. Cette démarche s'accompagne d'ateliers à destination des étudiants afin de les former à l'écosystème informationnel, à les sensibiliser notamment à leur identité numérique.

Dans les mois à venir, le SIGB[16] existant, qui est un outil conçu par les Geemiks en 1997, va être remplacé par l'application open-source *Koha*[17], tant au niveau du campus de Lille que des autres campus.

2.3.2 Public et missions

Les communautés du campus de Lille auxquelles s'adressent les Geemiks comprennent à la fois le public direct de l'école (étudiants, enseignants, chercheurs, collaborateurs et diplômés), mais également des experts, consultants, acteurs en région et cadres d'entreprise. Le travail des Geemiks consistent à[18] :

- à faciliter l'accessibilité aux ressources, à développer des moyens d'être là où sont les utilisateurs ;

[16] Système intégré de gestion de bibliothèque
http://fr.wikipedia.org/wiki/Syst%C3%A8me_int%C3%A9gr%C3%A9_de_gestion_de_biblioth%C3%A8que
[17] http://www.koha-fr.org/
[18] Interview d'Isabelle Brisset, blog B-R-ENT - le blog des managers 2.0, 10 mars 2009
http://b-r-ent.com/news/interview-exclusive-d-isabelle-brisset-pour-tout-savoir-sur-les-geemiks-sur-youontheweb

- à développer de nouveaux services pour favoriser l'échange et la contribution entre les différentes communautés ;
- à créer des produits à valeur ajoutée pour les communautés ;
- à former l'ensemble de son public et l'accompagner dans l'adoption de ces nouveaux modes de travail et d'apprentissage : 3 heures de formation assurées en début d'année, complétées par des ateliers tout au long de l'année aux meilleures méthodes et outils pour s'informer et se former ;
- à témoigner dans les réseaux de cette expérience et à animer un groupe thématique de retour d'expériences sur l'activité d'animatrice de communauté et sur les thèmes ingénierie des connaissances et de l'intelligence collective ;
- à animer un espace, la Fusée, qui soit le reflet de ces nouveaux territoires d'apprentissage, un espace augmenté qui mêle espace physique et espace virtuel, un lieu de « serendipité », de curiosité « où l'on puisse découvrir également ce que l'on ne cherchait pas » ;
- à animer plusieurs événements dans l'année autour de leurs thématiques, à partir de septembre 2010.

2.3.3 L'accès aux ressources

Le problème majeur du système actuel est de proposer un accès non centralisé aux ressources, ce qui rend celles-ci difficiles à interroger. Deux cas de figure se présentent à l'internaute qui arrive sur le portail pour effectuer une recherche. Dans le premier cas, pour un accès au contenu matériel de la médiathèque, l'utilisateur doit utiliser le champ unique situé en haut de la page, dans la colonne de gauche (on notera la présence d'un flux RSS qui permet de se tenir informé des nouveautés) :

Figure 1 : Interface pour requêter le moteur AquaBrowser

La page de résultats est proposée avec une interface sur trois colonnes (Figure 2) :

- un nuage de mots-clés qui permet de voir les termes apparentés ; cliquer sur l'un de ces termes permet de basculer sur une nouvelle recherche sur ce terme ;
- les résultats, présentés sous forme de fiche, avec un repère visuel précisant le type de document (ouvrage, usuel, thèse, etc.);
- une classification à facettes qui permet d'affiner la recherche.

Figure 2 : Interface de résultats du moteur AquaBrowser

Le second cas de figure qui se présente est l'accès aux ressources numériques. Les différents types de ressources sont proposés dans le menu de gauche (Figure 3).

Figure 3 : Menu d'accès aux ressources numériques

Chaque lien du menu renvoie vers une fiche décrivant les ressources proposées :

1. ***E-books :*** portail ScholarVox, de la société Cyberlibris ;
2. ***Catalogue des revues :*** service *A to Z* d'EBSCO, pour parcourir la liste des revues auxquelles la médiathèque est abonnée. Ce service, en offre d'essai, ne va pas être prolongé ;
3. ***Synthèses d'ouvrages*** (Figure 4) : service *Business Digest* qui propose des dossiers sur le management et la stratégie. L'abonnement à ce service vient d'être stoppé et ne sera pas reconduit ;

Figure 4 : Fiche de présentation du service *Business Digest*

4. ***Bases de données*** (Figure 5) : l'abonnement aux divers bases de la médiathèque : *Base Export, Business Digest, Business Source Complete, Dalloz, ExpertiseNet, Factiva, Jstor, Marketline, MintGlobal, Mémentis, Navis Integral, TerraEconomica* et *Xerfi700.*

Figure 5 : Page d'accès aux bases de données

5. ***Bases de données de test :*** Palgrave Dictionary of Economics, E-Journals : sommaires en ligne ;
6. ***Recherche orientée*** (Figure 6) : Carte sous forme d'arborescence qui présente les différentes bases en fonction des thématiques ;

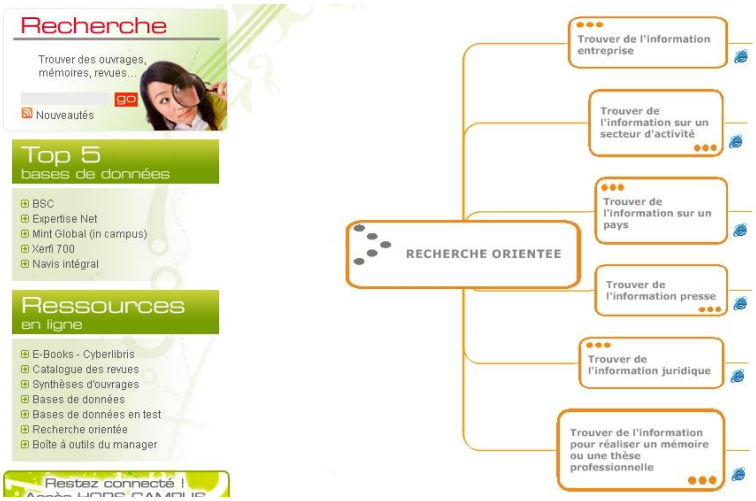

Figure 6 : Recherche orientée sous forme de carte arborescence

7. ***Boîte à outils du manager :*** il s'agit là aussi d'une carte arborescente proposant des liens vers des outils en ligne.

Ainsi, par exemple, un étudiant qui souhaite rechercher une information parue dans la presse choisira d'utiliser la base *Factiva*. Pour accéder à l'interface de recherche de cette base, il devra passer par les étapes suivantes :

1. Se rendre sur la page d'accueil de la médiathèque ;
2. Cliquer sur le lien « Bases de données » dans le menu ;
3. Cliquer sur le logo *Factiva* pour l'accès à la fiche de *Factiva* ;
4. Cliquer sur le lien approprié qui renvoie vers la page d'accueil de Factiva ;

Il est néanmoins possible de gagner une étape, à l'étape 2, en cliquant sur l'une des icônes jaunes (selon que l'on se trouve sur le campus ou hors-campus) (Figure 7). Ces liens permettent d'accéder à la page de recherche *Factiva*, sans passer par la fiche associée.

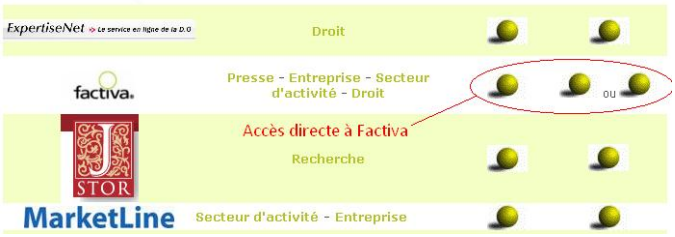

Figure 7 : Liens pour l'accès direct au site *Factiva*

A noter également que la page d'accueil de la médiathèque propose, dans le premier menu de la colonne de gauche, des liens directs vers les cinq bases les plus consultées :

Comme nous venons de le voir, l'accès aux ressources est hétérogène ; le parcours pour atteindre l'interface de recherche des différentes bases de données est généralement séquencé par trop de pages écran, ce qui est un réel frein à leur utilisation (Figure 8).

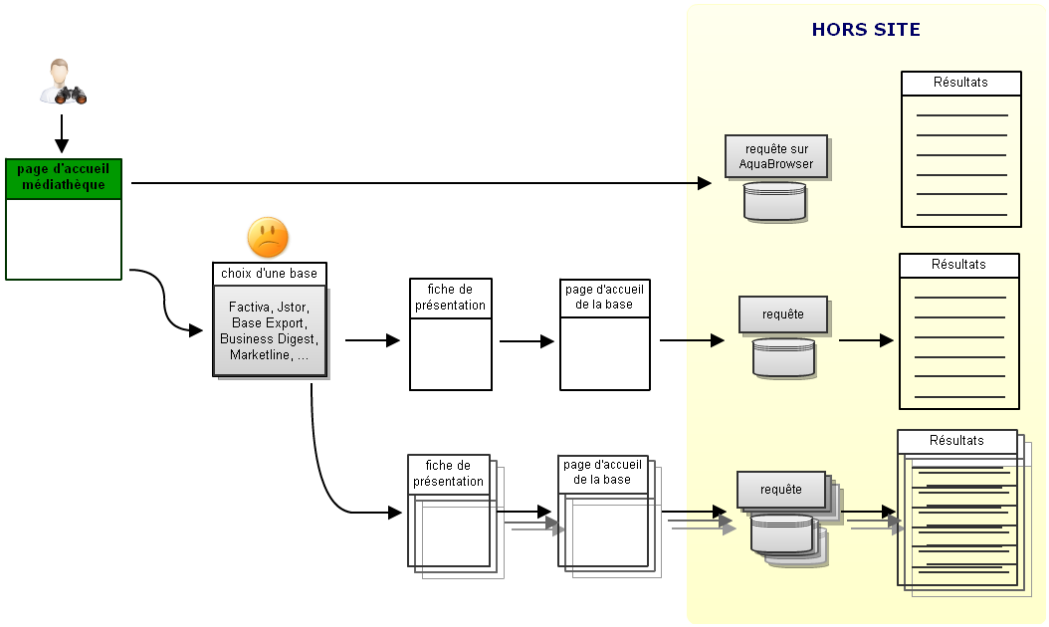

Figure 8 : Système actuel de recherche (multiples points d'entrée)

La règle des « trois clics » (Zaphiris et Mtei, 1997) stipule que, sur un site web, les utilisateurs ne doivent pas faire plus de trois « clics » pour trouver ce qu'ils cherchent, sinon ils peuvent manifester du mécontentement, voire quitter le site. L'étude a été remise en question en 2003 par l'équipe de l'User Interface Engineering[19] (UEI) pour aboutir à la conclusion que ce qui compte ce n'est pas le nombre de « clics », mais que les utilisateurs arrivent à leur fin. Pour Fred Cavazza[20], réduire le nombre de « clics » sur un site revient à augmenter la charge cognitive (car cela suppose d'augmenter le nombre de choix possibles par pages). Il propose donc de privilégier la simplicité d'utilisation en proposant aux utilisateurs des choix plus simples même si cela les amène à opérer plus de sélection avec le pointeur (souris).

Selon (Tallent, 2004), la ségrégation des types de documents sur des bases distinctes constituerait l'un des problèmes les plus frustrants pour les usagers qui ne comprennent généralement pas pourquoi il est impossible de retrouver des références d'articles de périodiques dans le catalogue de la bibliothèque. Ce qui tendrait à démontrer que les

[19] Joshua Porter, *Testing the Three-Click rule*, sur le site User Interface Engineering, 16 avril 2003 http://www.uie.com/Articles/three_click_rule.htm

[20] Frédéric Cavazza, *En finir avec la règle des 3 clics*, blog de Frédéric Cavazza, 28 novembre 2003 http://www.fredcavazza.net/2003/11/28/en-finir-avec-la-regle-des-3-clics/

utilisateurs réutilisent les sources qu'ils connaissent déjà et pour lesquels ils ont obtenus des résultats satisfaisants.

2.4 Le projet de Knowledge Center

2.4.1 Étude des besoins

Ce projet doit se concrétiser à terme sous la forme d'un portail unique d'accès à l'information afin de mettre la richesse et la diversité de l'offre au service de tous, tout en répondant aux besoins de chacun. Conçu pour la recherche multi-bases et multicritères, le portail doit proposer une interface d'accès qui facilite la visibilité et l'appropriation dans l'usage. Il doit bénéficier d'une approche qui mise sur la simplicité, qui supporte des règles et des standards, et présente une offre intégrée de l'imprimé et de l'électronique.

En plus de gérer les différents types de ressources (ouvrages, périodiques, thèses, …), ce moteur de recherche doit avoir une réelle compréhension des différents objets propres à certaines bases (tels que des études de marché, des bilans financiers, des informations sur les entreprises, etc.) et fournir un moyen d'accès spécifique et adapté à leur nature. Les ressources doivent être rendues accessibles plus largement, sous forme de services web (par le biais d'une interface pour l'accès depuis un téléphone portable, permettre l'intégration sous forme de « widget », etc.), afin de contribuer à leur utilisation en toute occasion et assurer la rentabilité des abonnements. En résumé, les fonctionnalités souhaitées pour ce nouveau moteur de recherche sont :

- bénéficier d'une gestion des rôles en lien avec le coffre d'authentification (notion de « package » selon le profil et les ressources disponibles) ;
- permettre la recherche thématique (par domaine et sur certaines sources) ;
- proposer l'accès au moteur de recherche via des interfaces variées (intégré sous forme de widget, intégré au navigateur, …), assurer la mobilité pour être « là où ils sont »;
- permettre l'affinage et la sauvegarde des résultats ;

- permettre la sauvegarde des documents trouvés pour un usage « hors ligne », lorsque cela est possible ;
- fournir des fonctionnalités sociales (mise en favoris, ajout de commentaires, envoi par e-mail, …) ;
- fournir des flux RSS personnalisés pour être alerté en fonction de certains critères ;
- prendre en compte l'aspect sémantique des métadonnées pour l'évolution future du moteur.

La Figure 9 illustre l'intérêt d'un tel outil : un seul point d'entrée (un champ de recherche simple) pour l'accès aux différentes ressources.

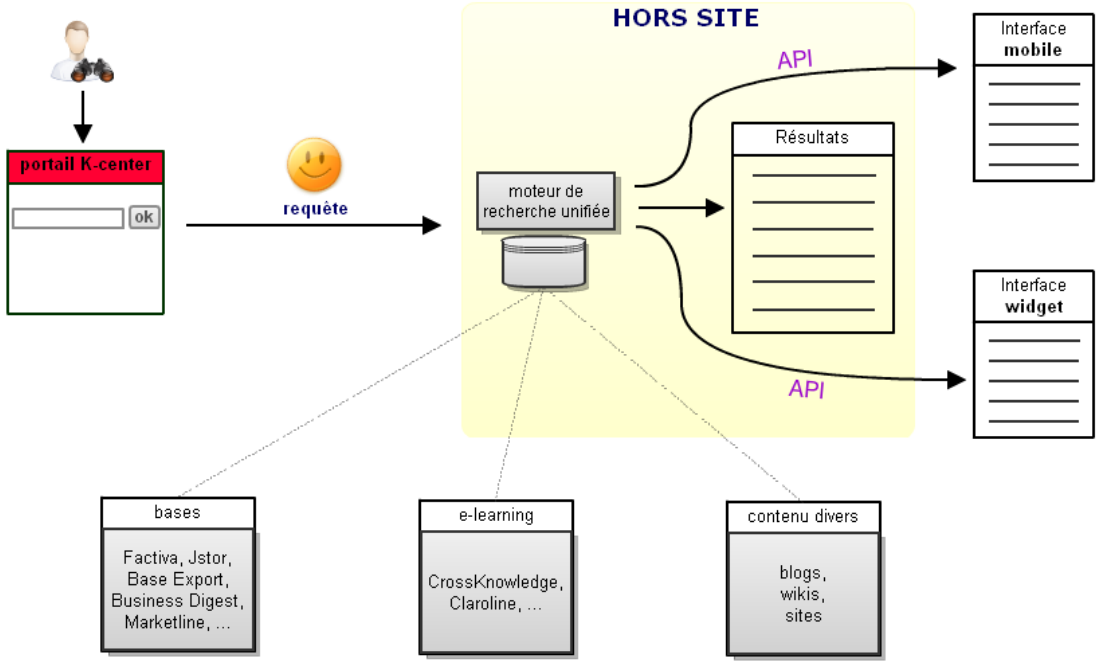

Figure 9 : Système de recherche unifiée (un seul point d'entrée)

Cette évolution se traduit par un glissement d'une conception intégrée des systèmes d'information vers une architecture de briques métiers modulables. Pour réussir le choix et la mise en place de ce portail, il est nécessaire de procéder par étape. Dans un premier temps, c'est la recherche relative au fonds documentaire de la médiathèque du campus de Lille qui a été étudiée.

2.4.2 Les ressources documentaires

Le fonds documentaire de la médiathèque de Lille est constitué de deux types de ressources : les imprimés et les publications numériques. Le catalogue des ouvrages et périodiques papier est exporté vers le moteur *AquaBrowser* tous les soirs. Ce procédé sera revu lors de la mise en place du nouveau SIGB (*Koha*) qui doit se faire en juillet.

Les abonnements à des offres numériques sont multiples et sujets à évolution (résiliation, offres d'essai, nouvelles bases issues de la fusion avec le CERAM de Nice). Aujourd'hui, cette offre numérique se compose comme suit :

- des livres électroniques (via *ScholarVox* de la société Cyberlibris) ;
- des bases de données (Base Export, Business Source Complete, Dalloz, ExpertiseNet (Jurisclasseur), Factiva, Jstor, MarketLine, MintGlobal, Mémentis, Navis Integral, TerraEconomica, Xerfi700).

Comme nous pouvons le constater, cette offre comporte un large panel de références (abonnement à plus d'une dizaine de bases) et représente un coût important (budget de 150000 euros). Dans une étude sur les bibliothèques numériques, (Papy et Leblond, 2009) notent que l'augmentation des ressources électroniques dans l'offre documentaire détourne de son objectif la formation générique des usagers à la recherche documentaire, en l'assujettissant à des formations instrumentales spécifiques aux interfaces de consultation dévolues aux ressources numériques. Si les modalités d'interrogation sont très similaires d'un environnement technique à l'autre, elles ne sont réellement exploitables que par une minorité. La grande majorité des usagers n'a qu'une utilisation très limitée des mécanismes sophistiqués d'interrogation, et sont rapidement confrontés aux problèmes de bruit et de silence documentaires (ibid). Dans l'étude de (Martin et Perret, 2009), seuls 25% des étudiants utilisaient la recherche avancée du moteur Google. (Brigitte Simonnot, 2009) a mené une enquête qui montre que le recours fréquent aux moteurs n'amène pas les étudiants à questionner les fondements d'une démarche informationnelle aboutie et n'améliore pas leurs connaissances info-documentaires.

3 LA RECHERCHE D'INFORMATION EN LIGNE

3.1 Les stratégies de recherche documentaire des étudiants

Une enquête menée par l'OCLC en 2005 (De rosa et al, 2005), nous apporte quelques chiffres sur l'usage des jeunes de la bibliothèque et des ressources d'information :

- 89% des lycéens utilisent des moteurs de recherche pour commencer une recherche d'information (alors que seulement 2% commencent par le site web d'une bibliothèque) ;
- 93% sont satisfaits ou très satisfaits de leur expérience globale d'utilisation d'un moteur de recherche (comparé à 84% dans le cas d'une recherche accompagnée par un bibliothécaire) ;
- les moteurs de recherche conviennent davantage au mode de vie des lycéens qu'une bibliothèque physique ou en ligne et cette adéquation est 'pratiquement parfaite' ;
- les lycéens utilisent encore la bibliothèque, mais ils l'utilisent moins (et il lisent moins) depuis qu'ils ont commencé à utiliser des outils de recherche sur internet ;
- les livres sont toujours la première caractéristique de la bibliothèque pour ce groupe, malgré les investissements massifs dans les ressources numériques auxquelles les étudiants sont pour la plupart peu habitués.

Les comportements, attitudes et préférences de la « génération Google[21] » ont fait l'objet d'une étude parue en 2008 (Rowlands et al, 2008), commandée par la British Library et le JISC[22]. Dans ce rapport, le graphique suivant (Figure 10) montre la valeur relative que les

[21] Expression qui fait référence à une génération de jeunes nés après 1993 et qui grandit dans un monde dominé par Internet. Le terme génération Y désigne les personnes nées entre la fin des années 1970 et le milieu des années 1990. L'expression « digital natives » est également utilisée pour pointer le fait que ces enfants ont grandi dans un monde où l'ordinateur personnel et l'Internet sont devenus de plus en plus accessibles.

[22] L'étude a été traduite en français en mars 2010.
http://www.educnet.education.fr/veille-education-numerique/mai-2010/le-comportement-informationnel-du-chercheur-du-futur

membres de la communauté universitaire attribuent à un ensemble de méthodes pour trouver des articles. Les différences liées à l'âge sont remarquables et laissent penser que le glissement de la bibliothèque physique vers la bibliothèque virtuelle va s'accélérer très rapidement, et que des outils comme *Google Scholar* vont de plus en plus être une menace réelle et présente pour la bibliothèque en tant qu'institution. (ibid)

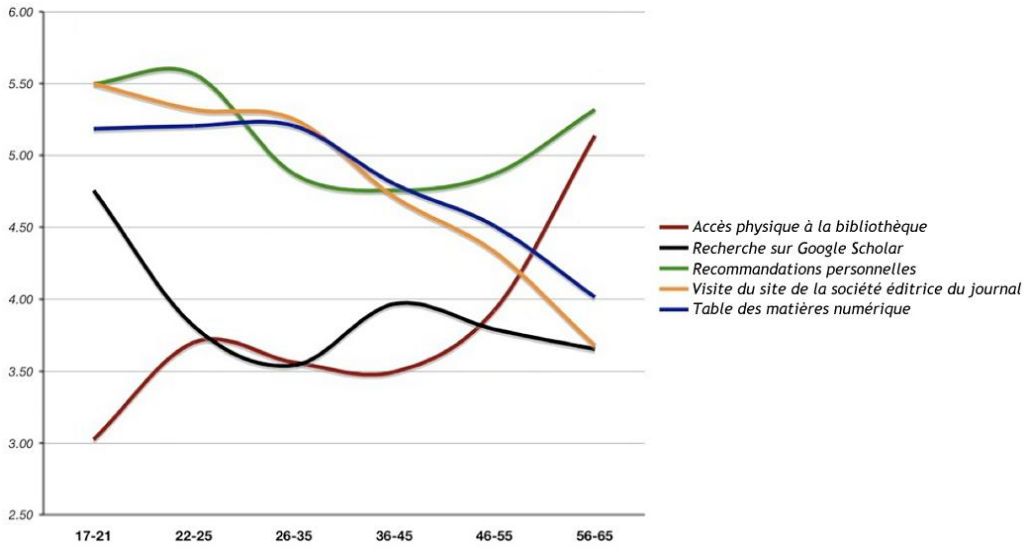

Figure 10 : Différences significatives des méthodes de découverte d'articles en fonction de l'âge

D'après l'étude menée par l'UCL (Nicholas et al, 2007), à partir des fichiers de logs[23] d'environ 500000 étudiants, les grands principes du comportement de recherche dans une bibliothèque numérique sont[24] :

- **La recherche d'information horizontale**.
  « Activité de forme écrémage, au cours de laquelle les gens regardent juste une ou deux pages d'un site scientifique et soudain « s'échappent » pour ne probablement plus jamais revenir. Les chiffres sont significatifs : environ 60% d'utilisateurs de journaux en ligne ne regardent pas plus de trois pages et une majorité (jusqu'à 65%) ne revient jamais. »

[23] L'analyse des fichiers logs pour la mesure d'audience sur Internet
http://fr.wikipedia.org/wiki/Mesure_d%27audience_sur_Internet#L.27analyse_des_fichiers_Logs
[24] Cité dans le rapport de Rowlands et al (2008) sur la génération Google.

- **La navigation**.

  « Dans les bibliothèques virtuelles, les gens passent beaucoup de temps à simplement chercher leur chemin : en fait ils passent autant de temps à chercher qu'à réellement lire ce qu'ils ont trouvé. »

- **Les lecteurs**.

  « Le temps moyen que les usagers passent sur des livres numériques et les sites de journaux en ligne sont très courts : typiquement quatre et huit minutes respectivement. Il est évident que les utilisateurs ne lisent pas en ligne au sens classique du terme, et de fait il y a des signes tangibles que de nouvelles formes de lecture émergent chez les utilisateurs qui naviguent horizontalement à travers les titres, les pages de contenu et les résumés à la recherche de réponses rapides. Il semble presque qu'ils vont en ligne pour éviter de lire au sens traditionnel du terme. »

- **Le syndrome de l'écureuil**.

  « Les recherches montrent que ses usagers collectent des contenus par téléchargements, surtout quand l'offre est gratuite. Malgré cela et les temps de connexion très courts que l'on peut attester, personne ne sait jusqu'à quel point ces téléchargements sont réellement lus. »

- **La diversité des chercheurs d'information**.

  « L'analyse des logs révèle que le comportement de l'utilisateur est très varié : situation géographique, sexe, type d'université et statut sont tous de forts marqueurs des groupes. Il n'y a pas de modèle universel. »

- **Vérification des chercheurs d'information**.

  « Les utilisateurs évaluent l'autorité et la pertinence de l'information par eux-mêmes en quelques secondes en piochant dans différents sites et en croisant les informations, tout en s'appuyant sur les grands noms commerciaux (par exemple Google). »

Selon (Nielsen, 2006), une fois sur deux, l'utilisateur ne trouve pas ce qu'il cherche. Une étude de (Ihadjadene, 1999) a permis d'identifier les raisons des échecs sur l'interrogation de

deux catalogues en ligne (ENSSIB et Lyon 2). La principale cause est due au fait que les utilisateurs effectuent des erreurs lors de l'écriture des requêtes. Les autres causes identifiées sont l'usage de termes en langue étrangère, la précision des requêtes et la difficulté d'usage des opérateurs booléens ou des troncatures. (Maisonneuve, 2003) fait remarquer que 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.

D'après l'étude menée par (Spink et al, 2001), le nombre moyen de mots par requête s'élève à 2,4 et environ 27% des demandes se limitent à un seul terme. Par conséquent, leurs interrogations aboutissent souvent à des surcharges d'informations. Seulement un usager sur deux tente de réduire le nombre de réponses en ajoutant souvent un terme à l'équation d'origine (Favier et Ihadjadene, 2000). Les reformulations consistent généralement dans la répétition de la requête initiale, la substitution d'un ou plusieurs termes d'une requête ou l'ajout ou le retrait de mots de la requête (Bruza et Dennis, 1997). Enfin, pour (Marchionini, 1995), l'expertise du domaine est un facteur important pour la performance de recherche d'information. En effet, les experts formulent mieux leurs requêtes et savent mieux discerner les résultats pertinents. Mais pour (Hsieh-Yee, 1993), les résultats des recherches montrent que la connaissance du domaine est un facteur faible qui ne joue pas sur le déroulement des interrogations d'un catalogue en ligne.

3.2 Les systèmes de recherche d'information

Un système de recherche d'information (SRI) est un ensemble de modèles et de processus qui vise à retourner à l'usager un ensemble de documents. L'objectif d'un SRI est d'optimiser le nombre de documents pertinents retournés (Bonnel, 2008). (Morizio, 2006) rappelle que d'un type de système plutôt générateur de silence, nous avons basculés vers un type de système d'avantage générateur de bruit, suite à l'explosion documentaire et la diversification des modes d'accès à l'information et aux documents. Les systèmes actuelles impliquent ainsi pour l'usager de savoir opérer des sélections dans les corpus de réponses proposés.

La recherche en ligne ne cesse de croître et de manière fulgurante, si l'on en croit les derniers chiffres de ComScore[25], l'un des leaders mondiaux de la mesure d'audience d'Internet. 131 milliards de requêtes ont été effectuées sur Internet en décembre 2009, soit 46% de plus qu'en décembre 2008. En France[26], un internaute effectue en moyenne 2,6 recherches par jour sur Internet. Ces chiffres montrent l'importance des moteurs de recherche pour les internautes.

(Morville, 2007) définit quatre grands types de besoin en information (Figure 11) :

- *la recherche exhaustive* : lorsque l'utilisateur veut savoir tout ce qui est disponible sur un sujet précis ;
- *la recherche d'une information connue* : lorsque l'utilisateur sait ce qu'il cherche, comme un numéro de téléphone par exemple ;
- *la recherche exploratoire* : lorsque l'utilisateur n'est pas sûr de ce qu'il cherche. Il ne s'attend pas à trouver la « bonne réponse », il est satisfait de trouver quelques informations pertinentes qu'il pourra utiliser comme base de départ pour une prochaine investigation ;
- *la recherche d'une information trouvée antérieurement.*

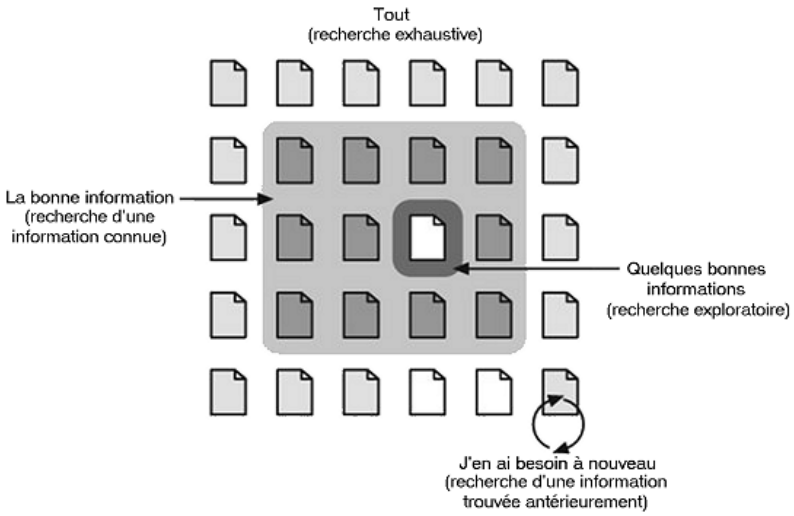

Figure 11 : Les quatre grands types de besoin en information (d'après Morville, 2007)

[25] ComScore. *La recherche en ligne a progressé de 46% en un an.*, 25 janvier 2010
http://www.zdnet.fr/actualites/informatique/0,39040745,39712445,00.htm
[26] Médiamétrie. *Médiamétrie/NetRatings annonce un nouveau service de mesure des comportements de recherche en ligne des internautes*, avril 2009
http://www.mediametrie.fr/internet/communiques/telecharger.php?f=9a1158154dfa42caddbd0694a4e9bdc8

(Balmisse, 2008) identifie quant à lui cinq types de recherche les plus couramment rencontrées, en fonction de l'information recherchée et de l'expression des besoins (Figure 12).

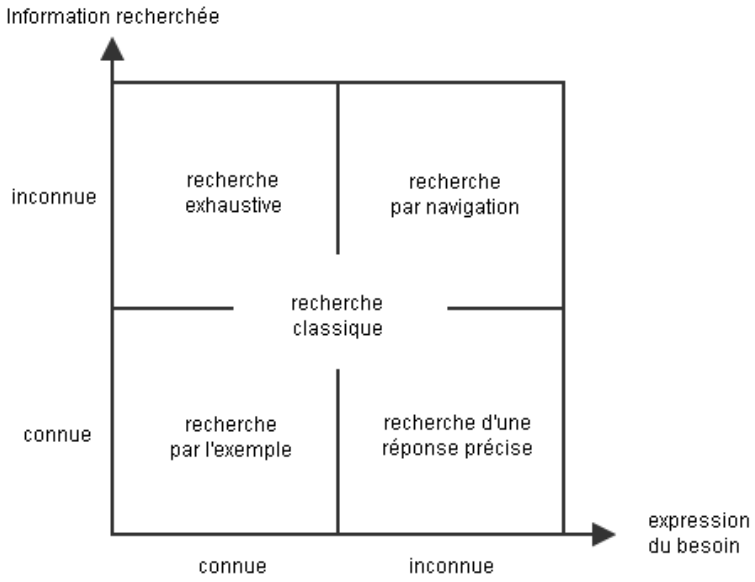

Figure 12 : Les différents modes et types de recherche, d'après Gilles Balmisse (2008)

La recherche d'information peut être découpée en plusieurs étapes que l'on peut illustrer de la manière suivante (Figure 13) :

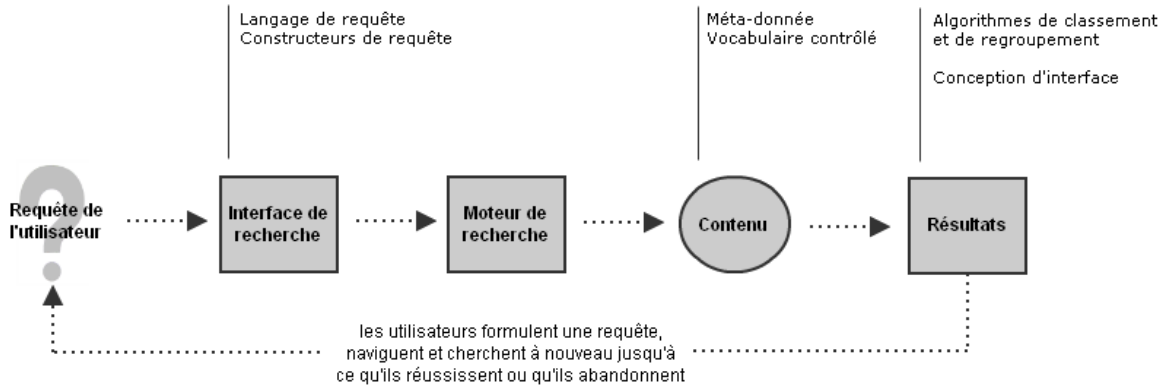

Figure 13 : L'anatomie de base d'un système de recherche, d'après Morville (2007)

3.3 Aspects cognitifs de la recherche d'information

La recherche d'information ne se réduit pas à un processus mécanisable dont la maîtrise assurerait l'efficacité. Elle fait entrer en jeu des dimensions cognitives et affectives dont la prise en compte est indispensable pour en appréhender la complexité (Morizio, 2006). Avant d'interroger un moteur de recherche, l'utilisateur construit une représentation mentale de son besoin d'information (Chieze, 2006). L'étude des comportements des utilisateurs autour des SRI a mis en évidence plusieurs modèles de processus parmi lesquels :

- **le modèle en « croissance de perles » (*pearl growing*) de (Hawkins et Wagers, 1982)**. Dans ce modèle, l'utilisateur débute sa recherche avec un ou plusieurs documents correspondant exactement à ses besoins. Il souhaite alors récupérer d'autres documents similaires à l'un de ceux dont il dispose.

- **le modèle de « cueillette de baies » (*berry picking*) de (Bates, 1989)**. Dans ce modèle, l'utilisateur recueille l'information de façon discontinue et dispersée, une par une. Il reformule sa requête au fur et à mesure qu'il découvre des informations qui répondent à sa demande et en fonction de celles qui sont accessibles via le système.

- **le modèle de la course d'orientation de (O'Day et Jeffries, 1993)** : les utilisateurs se servent des résultats obtenus pour déterminer leur prochaine recherche ; ils cherchent avant tout à élaborer une vue d'ensemble de la recherche.

- **le modèle de représentation en pivot de (Marchionini, 1995)** : les usagers modifient leur stratégie de recherche en fonction du contexte pour maximiser leur activité.

D'après (Morizio, 2006), le besoin d'information tend à se modifier en fonction des documents trouvés. Le volant final de documents considérés comme pertinents sera souvent bien différent de celui qui avait fait partie de la représentation mentale de l'utilisateur au départ. Par ailleurs, la recherche par sujet, qui représente la grande majorité des recherches d'information, demande des capacités d'abstraction et de conceptualisation qui la rende difficile pour beaucoup d'utilisateurs. Pour obtenir une réponse, il doit non seulement préciser son besoin d'information, mais le préciser de façon qui convienne au système. Pour (Navarro-Prieto et al, 1999), les utilisateurs expérimentés planifient à l'avance leur comportement de

recherche basé sur leur connaissance du Web, tandis que les utilisateurs inexpérimentés planifient difficilement et sont davantage guidés par les représentations externes ; par ce qu'ils observent sur le moment sur l'écran.

3.4 Visualisation et interfaces de recherche

3.4.1 Le rôle de l'ergonomie pour les bibliothèques numériques

L'ergonomie est *« la discipline scientifique qui vise la compréhension fondamentale des interactions entre les humains et les autres composantes d'un système, et la mise en œuvre dans la conception de théories, de principes, de méthodes et de données pertinentes afin d'améliorer le bien-être des hommes et l'efficacité globale des systèmes »*[27]. L'ergonomie des logiciels porte sur l'aspect cognitif de la RI[28], pour un traitement et une représentation des informations conformes aux attentes et capacités des utilisateurs (Calvary, 2002). (Bastien et Scapin, 1993) retiennent huit critères ergonomiques pour l'évaluation d'interfaces utilisateurs :

- **le guidage de l'utilisateur.**

« Le guidage est l'ensemble des moyens mis en œuvre pour conseiller, orienter, informer et conduire l'utilisateur lors de ses interactions avec l'ordinateur (messages, alarmes, labels, etc.) y compris dans ses aspects lexicaux ». Quatre sous-critères y participent : l'incitation, le groupement/distinction entre items, le retour d'information immédiat et la lisibilité ;

- **la charge de travail.**

« La charge de travail concerne l'ensemble des éléments de l'interface qui ont un rôle dans la réduction de la charge perceptive ou mnésique des utilisateurs et dans l'augmentation de l'efficacité du dialogue ». Deux sous-critères y contribuent : la brièveté et la densité informationnelle ;

[27] définition de référence proposée par l'International Ergonomics Association (IEA) en 2000
[28] RI pour Recherche d'Information

- **le contrôle explicite.**

« Le contrôle explicite concerne à la fois la prise en compte par le système des actions explicites des utilisateurs et le contrôle qu'ont les utilisateurs sur le traitement de leurs actions ». Deux sous-critères y participent : les actions explicites et le contrôle utilisateur ;

- **l'adaptabilité.**

« L'adaptabilité d'un système concerne sa capacité à réagir selon le contexte et selon les besoins et préférences des utilisateurs ». Deux sous-critères y participent : flexibilité et prise en compte de l'expérience de l'utilisateur ;

- **la gestion des erreurs.**

« La gestion des erreurs concerne tous les moyens permettant d'une part d'éviter ou de réduire les erreurs, et d'autre part de les corriger lorsqu'elles surviennent ». Trois sous-critères y participent : la protection contre les erreurs, la qualité des messages d'erreur et la correction des erreurs ;

- **l'homogénéité/cohérence.**

« L'homogénéité/cohérence se réfère à la façon avec laquelle les choix de conception de l'interface (codes, dénominations, formats, procédures, etc.) sont conservés pour des contextes identiques et sont différents pour des contextes différents » ;

- **la signifiance des codes et dénominations.**

« La signifiance des codes et dénominations concerne l'adéquation entre l'objet ou l'information affichée ou entrée, et son référent. Des codes et dénominations 'signifiants' disposent d'une relation sémantique forte avec leur référent » ;

- **la compatibilité**.

« La compatibilité se réfère à l'accord pouvant exister entre les caractéristiques des utilisateurs (mémoire, perceptions, habitudes, compétences, âge, attentes, etc.) et des tâches, d'une part, et l'organisation des sorties, des entrées et du dialogue d'une application donnée, d'autre part. De plus, la compatibilité concerne également le degré de similitude entre divers environnements ou applications ».

L'acceptation d'un système se réfère à son adéquation vis-à-vis des besoins et exigences des utilisateurs et autres interlocuteurs (Calvary, 2002). (Nielsen, 1993) propose ainsi un modèle des attributs de l'acceptabilité d'un système (Figure 14).

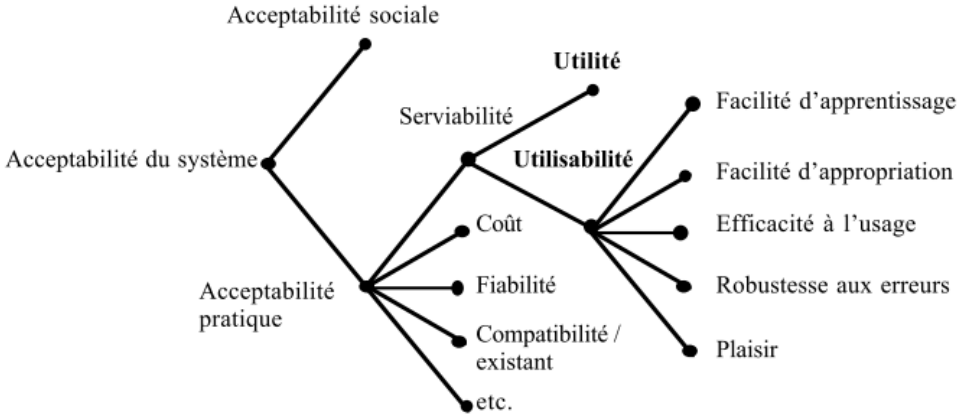

Figure 14 : Modèle des attributs de l'acceptabilité d'un système[29]

3.4.2 Exemples d'interfaces pour l'accès et la navigation dans les catalogues en ligne

(Chevalier, 2002) propose de classer les techniques de visualisation pour la recherche d'information de la manière suivante :

- visualisation des attributs des documents ;
  - visualisation de la répartition des termes de la requête ;
  - visualisation d'attributs prédéfinis ;
  - visualisation d'attributs définis par l'utilisateur ;

- visualisation des relations inter-documents ;
  - visualisation des documents ;
  - visualisation des classes ;

[29] Schéma tiré de Calvari (2002) pour la traduction française.

Pour la RI, la visualisation peut être utile pour réduire l'effort cognitif de l'utilisateur, l'aider dans l'orientation et améliorer certaines tâches parmi lesquelles (Hascoët, 2004) :

- l'exploration rapide d'ensembles d'informations inconnues ;
- la mise en évidence de relations et de structures dans les informations ;
- la mise en évidence de chemin d'accès à des informations pertinentes ;
- la classification interactive des informations ;

Parmi les méthodes de visualisation de l'information qui sont apparues ces vingt dernières années, (Fekete, 2008) retient pour les bibliothèques numériques :

- **les techniques visant à offrir une vue globale de l'espace d'information.**

« Elles fournissent des vues d'ensemble destinées à faire émerger des structures ou des comportements significatifs, des singularités ou des ressemblances qui vont aider l'utilisateur à trouver ce qui l'intéresse dans la « masse » des informations disponibles. »

- **les techniques focus + contexte et les métaphores tridimensionnelles.**

« Elles permettent de représenter le contexte d'un point local, c'est-à-dire les éléments qui entourent spatialement l'objet d'intérêt courant (ou qui lui sont liés par une relation donnée) sous une forme dégradée, mais néanmoins éclairante. Ce sont des représentations à deux niveaux d'échelle qui sont destinées à mettre en lumière les détails de quelques objets d'intérêt tout en conservant les relations entre ces objets et leur environnement. »

- **les représentations à niveau de granularité variable, appelées multi-échelle.**

« Elles sont généralement basées sur le concept de zoom sémantique, ce qui signifie que la notion d'échelle n'est pas purement géométrique mais que les représentations et leur niveau de détail varient avec le niveau d'échelle. Ces techniques permettent de passer continûment d'une vue globale à une vue détaillée et d'établir ainsi un lien cognitif entre les détails et la vue d'ensemble. »

Le parcours des utilisateurs face aux ressources mises à leur disposition est d'autant plus constructif que la structure associative de l'ensemble apparaît clairement (Chauvin, 2007). Dans le contexte des bibliothèques numériques, les techniques de visualisation sont particulièrement utiles pour la navigation au sein des collections multimédias. D'après l'étude de (Lippincott, 2005), les élèves du secondaire s'appuient fortement sur les informations affichées sous forme graphique sur les pages Web et s'appuient souvent sur des graphiques et des indices visuels pour interpréter la pertinence de ces pages. De son côté, (Sebrechts, 1999) montre que l'utilisateur est plus sensible aux représentations graphiques qu'à une représentation textuelle et que pour une même tâche la représentation 3D permet d'obtenir globalement de meilleurs résultats. (Keim et Kriegel, 1995) expliquent, pour leur part, que la coloration a un fort impact sur l'interprétation des résultats.

Les grands noms du web commencent à implémenter différentes solutions pour aider visuellement les internautes. Google fourni par exemple la fonctionnalité « roue magique » qui propose à l'internaute des termes apparentés à sa requête en cours. Il est possible de cliquer sur une recherche donnée pour afficher la roue des termes associés (Figure 15). Google propose également *Image Swirl*[30] pour la recherche d'images (Figure 16). *Ask Ken*[31] (Figure 17) et *Thinkpedia*[32] fournissent des interfaces quelque peu semblables pour naviguer dans les informations issues de Wikipedia.

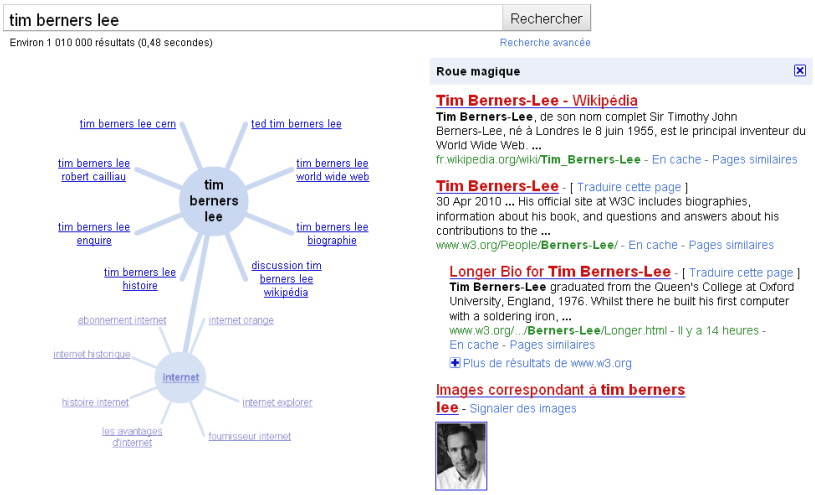

Figure 15 - La fonctionnalité "Roue magique" de Google

[30] http://image-swirl.googlelabs.com/
[31] http://askken.heroku.com/
[32] http://thinkpedia.cs.auckland.ac.nz/

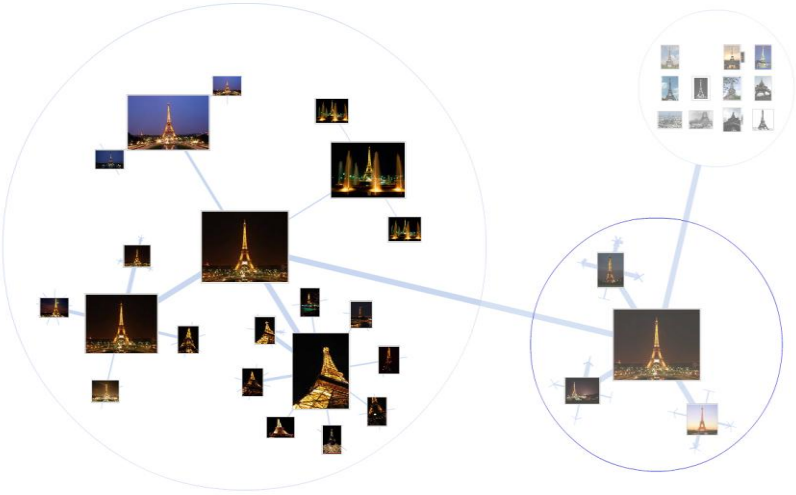

Figure 16 : La recherche d'images avec Google Image Swirl

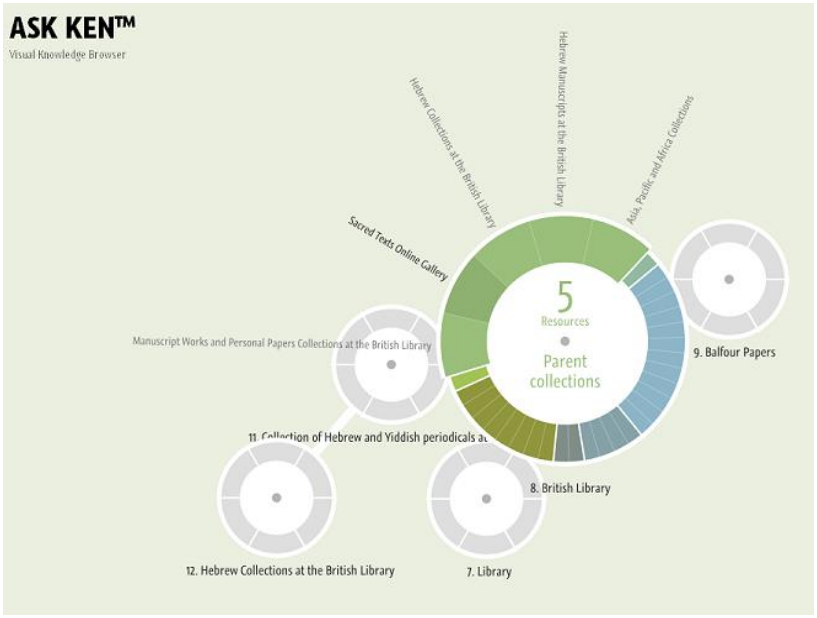

Figure 17 : Interface de navigation de terme à terme, avec conservation des traces de navigation

Dans la lignée du système de nuage de mots d'*AquaBrowser*, la bibliothèque municipale de Lille propose un mode de recherche « par constellation »[33] qui, à partir d'un terme de recherche donné, permet de voir les mots liés au terme en cours (Figure 18).

[33] http://portail.bibliotheque.bm-lille.fr/Typo3/index.php?id=zones-lille&fn=constellation2

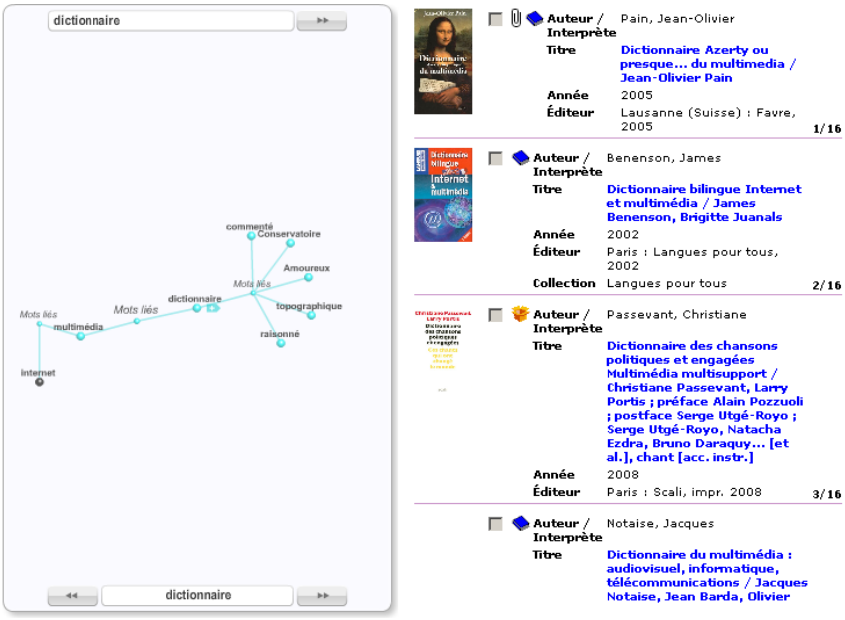

Figure 18 : Recherche par constellations sur le site de la bibliothèque municipale de Lille

Eyeplorer.com propose un moteur visuel original qui permet d'étudier les liens entre les différents articles tirés de Wikipédia. Les termes associés à la recherche de l'utilisateur sont affichés sous la forme d'une roue colorée (Figure 19). Sélectionner un terme et le placer au centre a pour effet d'affiner la recherche. Un survol sur un terme de la carte permet d'obtenir un descriptif. Dans un style proche des cartes heuristiques, *Crystal Mapping*[34] est un service de cartographie en ligne qui affiche les catégories dans des cercles concentriques qui sont subdivisés en segments de couleur distincte (Figure 20). Citons également le méta-moteur open-source *Carrot*[2] [35] qui présente les résultats accompagnés d'un arbre hyperbolique (Figure 21). En parallèle, *Wikipedia Roll*[36] propose de parcourir de manière dynamique et globale les contenus de l'encyclopédie collaborative avec une organisation du contenu par rotonde (Figure 22). Signalons enfin le projet KmP[37] (Knowledge Management Platform) d'@CTIS-Ingénierie qui possède une interface en demi-cercle utilisée pour la cartographie des acteurs, des compétences, des collaborations et des projets des clusters et pôles de compétitivité français (Figure 23).

[34] http://visualmapper.blogspot.com/2010/03/crystal-mapping-clear-thinking.html
[35] http://search.carrot2.org/stable/search
[36] http://api-exploration.net/mashups/wikipedia-roll/
[37] http://www.actis-ingenierie.com/kmp.htm

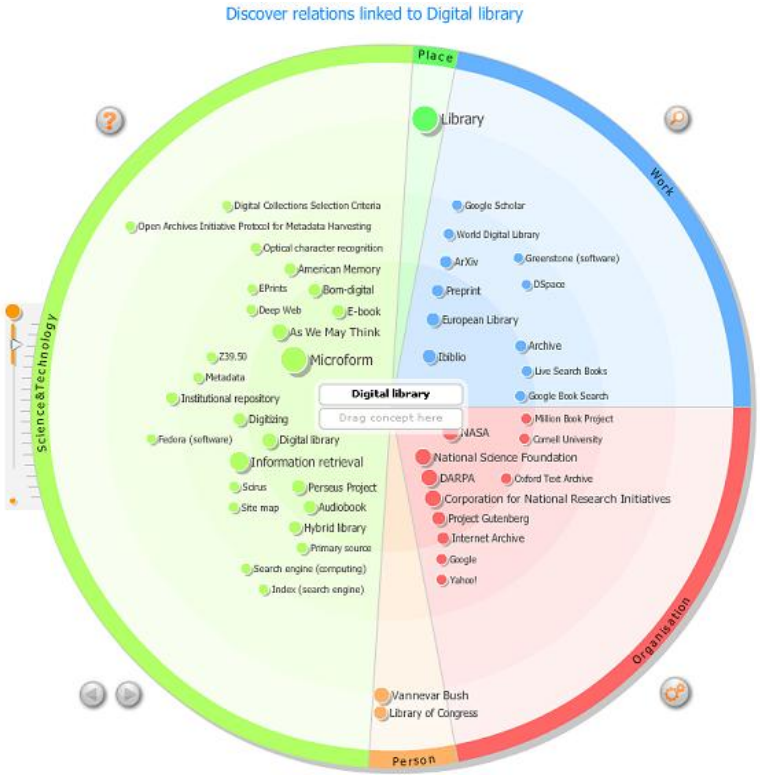

Figure 19 : L'interface du moteur EyePlorer

Figure 20 : Découpage circulaire de l'information avec Crystal Mapping

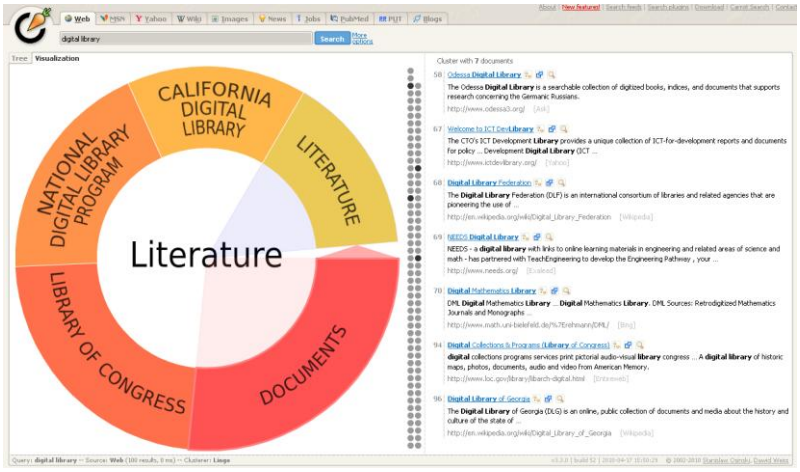

Figure 21 : Carrot2 et ses clusters sous forme d'arbre hyberbolique

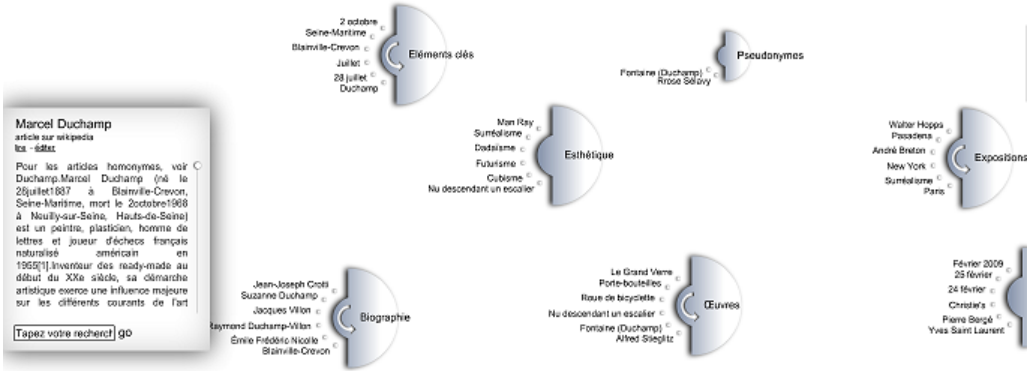

Figure 22 : Wikipedia Roll

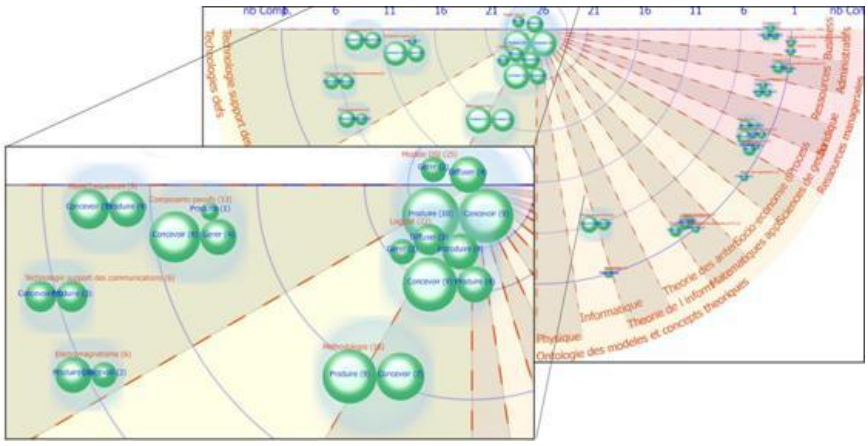

Figure 23 : Une visualisation en cluster au sein du portail KmP

A noter que les principaux moteurs innovants en terme d'interface qui étaient apparus ces dernières années (*Kartoo, Grokker, Kooltorch, Mugurdy, Mex-Search, ...*) ont désormais

disparus. *KartOOvisu*, de la société Kartoo, était certainement le plus intéressant dans sa façon de cartographier les différents sous-domaines d'une recherche. Basés sur les diagrammes de Voronoï[38], les résultats de recherche étaient accompagnés d'un affichage de mots-clés dont la superficie représentait la fréquence d'apparition de ce mot dans les résultats (Figure 24).

Figure 24 : Interface du moteur KartOOvisu

De son côté, *Grokker*[39] proposait un méta-moteur pour le web (ainsi qu'une version hors-ligne pour rechercher le contenu de son ordinateur) qui affichait les résultats en les divisant en ensembles et sous-ensembles circulaires (Figure 25). Les documents étaient représentés par des carrés plus ou moins grands. Le système a était utilisé notamment par *EBSCOhost*[40], ainsi que par la bibliothèque de Stanford[41] jusqu'en 2009, date à laquelle la société Groxis, créatrice du produit, a cessé son activité. Dans le cas de la bibliothèque de Stanford, il s'agissait plus précisément d'une application que l'utilisateur devait installer sur son poste de travail et qui permettait d'accéder aux contenus des différents abonnements de la

[38] http://fr.wikipedia.org/wiki/Diagramme_de_Vorono%C3%AF
[39] http://en.wikipedia.org/wiki/Grokker
[40] On trouvera des captures d'écran à cette adresse :
http://www.slideshare.net/mab992/information-visualization-using-visual-search-methods-272708/6
[41] http://library.stanford.edu/about_sulair/special_projects/stanford_grokker.html

bibliothèque, ainsi qu'aux résultats du web[42]. Le site de l'université précise qu'en 2005, l'application avait déjà été téléchargée plus de 2000 fois.

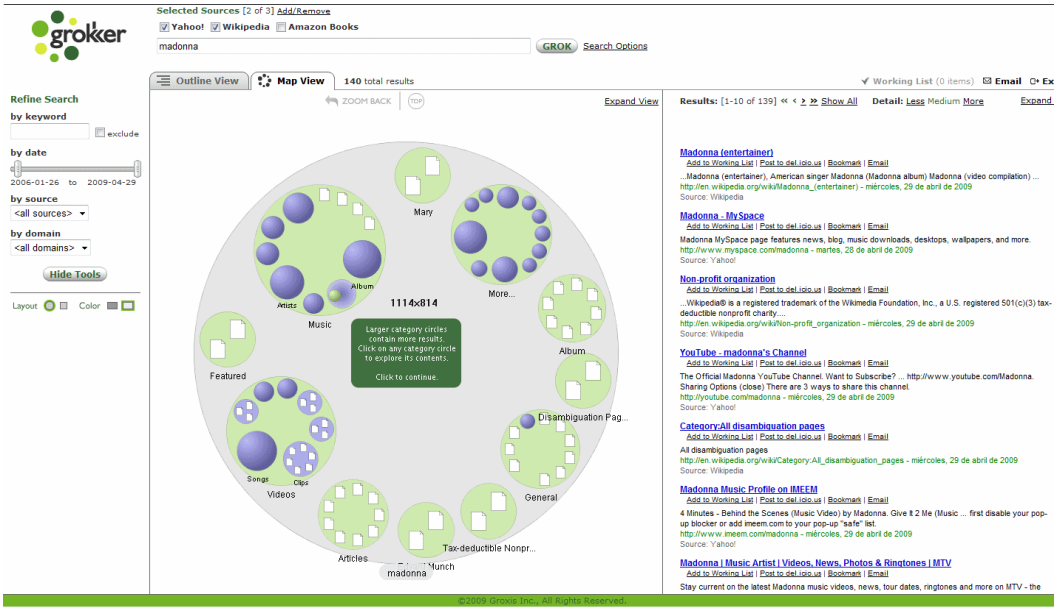

Figure 25 : Grokker : technique de représentation par pavage de cercles

Le projet *Elastic lists*[43] de Moritz Stefaner permet à partir d'une page de résultats de filtrer les informations selon les facettes qui sont proposées, afin d'affiner la recherche. Cette interface permet de naviguer dans des espaces riches et multi-dimensionnels en seulement quelques clics de souris (Figure 26).

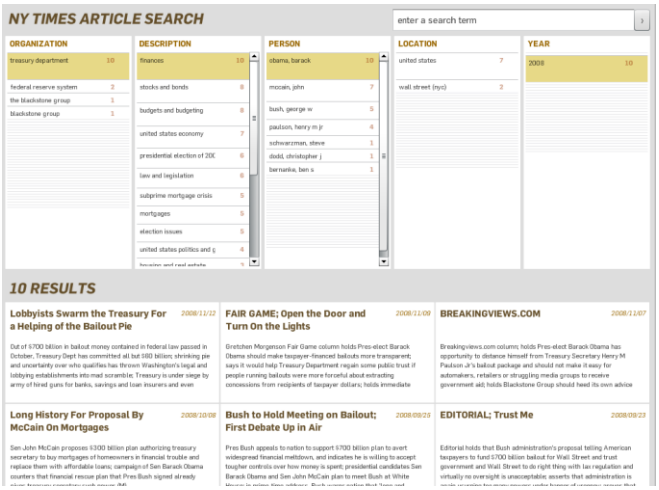

Figure 26 : Navigation à facettes dans les articles du New-York Times avec Elastic Lists

[42] http://library.stanford.edu/about_sulair/special_projects/stanford_grokker_faqs.html
[43] http://moritz.stefaner.eu/projects/elastic-lists/

(Stefaner, 2008) est également à l'origine du système de navigation du projet MACE[44] (Metadata for Architectural Contents in Europe) de la communauté européenne. Ce projet indexe plus de 600000 documents européens sur l'architecture. Plusieurs systèmes de navigation sont proposés : filtrage par facettes (avec *Elastic Lists*), par classification (arbre radial interactif, Figure 27), par compétence (filtrage par tableau), par lieu géographique (carte Google Maps) et par recherche sociale (nuage de mots-clés).

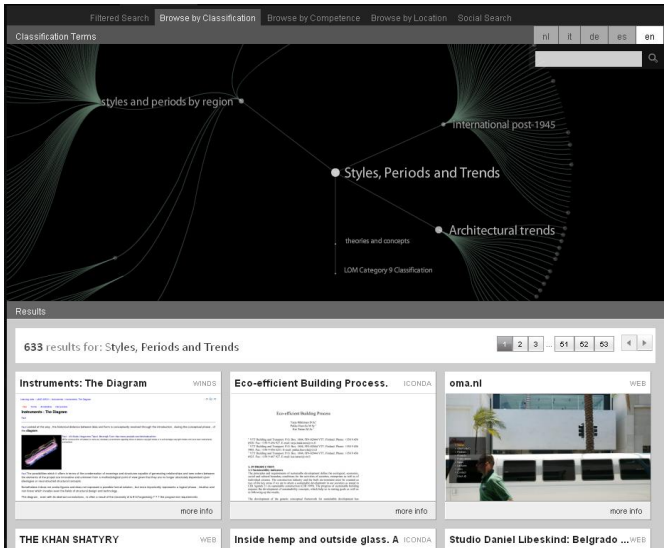

Figure 27 : Système de navigation dans les articles du projet MACE

Pour affiner la recherche, *EBSCO Host*[45] propose sur son portail un mode « recherche visuelle » qui propose des facettes en colonnes (Figure 28) à la manière d'*Elastic Lists*. Dans un premier temps, l'utilisateur se trouve face à une seule colonne qui propose une sélection de mots clés. En activant un mot, une seconde colonne apparaît proposant de nouveaux mots-clés, avec un niveau de granularité plus fin, et ainsi de suite.

[44] http://portal.mace-project.eu/
[45] http://www.ebscohost.com/

Figure 28 : L'interface de recherche visuelle d'EBSCO Host

Les nouveaux moteurs de recherche d'actualités, tels que *Evri*[46] et *SiloBreaker*[47], effectuent une analyse sémantique des contenus afin d'identifier des entités nommées[48]. Cela permet ensuite de proposer des options d'aide à la recherche (Figure 29 et Figure 30).

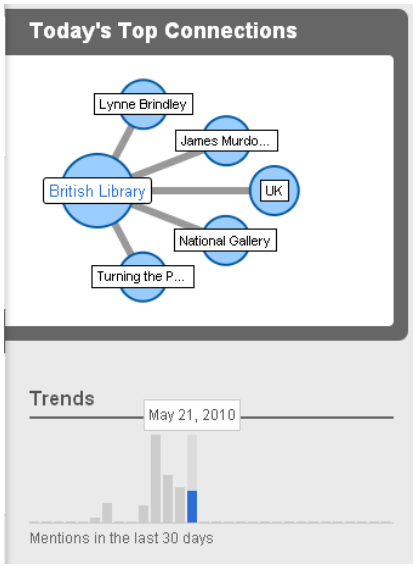

Figure 29 : Evri, graphe des connections et diagramme des tendances

[46] http://www.evri.com/
[47] http://www.silobreaker.com/
[48] Ce terme désigne un élément suffisamment unique et identifiable pour recevoir un nom qui lui est propre (noms d'individus et d'organisations, de lieux, d'expressions temporelles et numériques, ...).

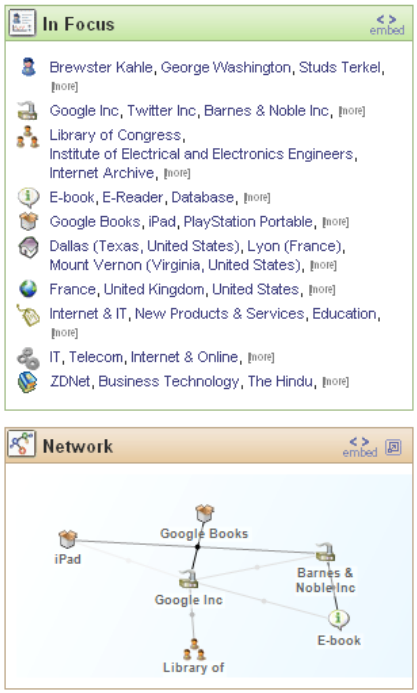

Figure 30 : Deux fonctionnalités de SiloBreaker : filtrage par facettes avec icônes et graphe de relations

Sinequa[49] propose une solution similaire pour les entreprises. On peut en voir une implémentation sur le site des Echos[50] (Figure 31).

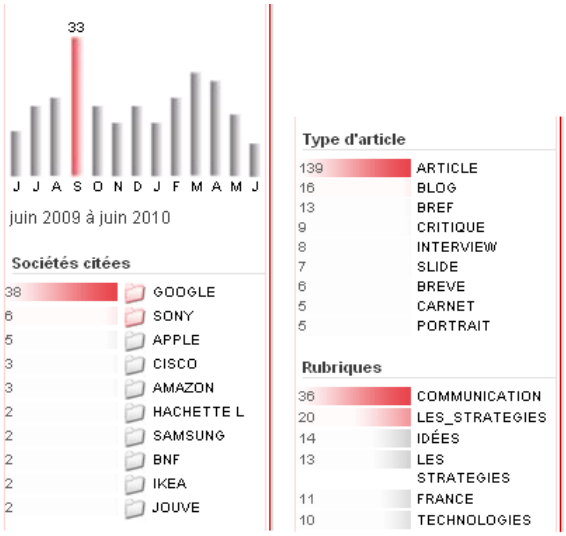

Figure 31: Indicateurs visuels, placés dans la colonne de droite, sur le moteur des Echos

[49] http://www.sinequa.com/
[50] http://recherche.lesechos.fr/

Le recours aux graphes pour la navigation est une tendance qui se développe autour des projets d'interfaces. On peut citer par exemple les projets *Touchgraph Google Browser*[51], *Constellations*[52] du français Exalead (Figure 32) et Cluuz[53].

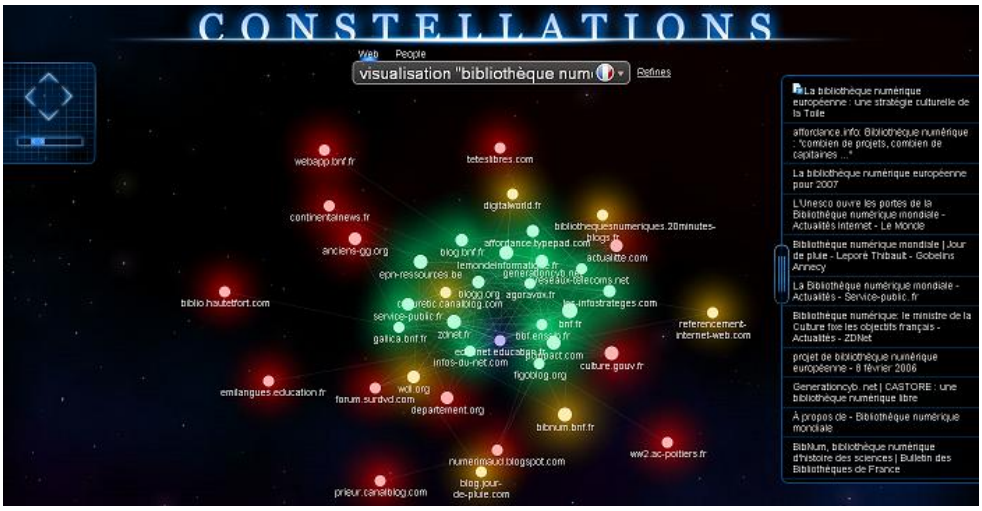

Figure 32 : Constellations, projet d'Exalead

Inspiré du système de recommandation d'Amazon, la bibliothèque de l'Université d'État Caroline du Nord a développé le *Virtual Shelf Index*[54], un système de navigation qui permet de parcourir le catalogue de la bibliothèque virtuelle, en offrant la possibilité de faire défiler les ouvrages en lien avec un livre donné (Figure 33). En recherche d'images citons également les moteurs *oSkope*[55] (mur de photos), *CoolIris*[56] (mur 3D), *Search Cube*[57] (cube 3D) et *Google Image Swirl* (graphe), cité plus haut.

[51] http://www.touchgraph.com/TGGoogleBrowser.html
[52] http://demos.labs.exalead.com/constellations/
[53] http://cluuz.com/
[54] http://www.lib.ncsu.edu/dli/projects/virtualshelfindex/
[55] http://www.oskope.com/
[56] http://www.cooliris.com/
[57] http://www.search-cube.com/

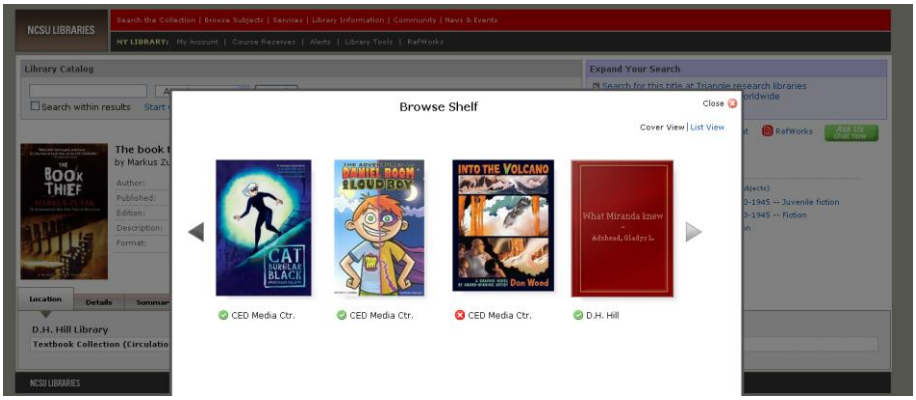

Figure 33 : Le projet *Virtual Shelf* Index de la bibliothèque de l'Université d'État Caroline du Nord

3.4.3 Le recours aux langages de pictogrammes et aux schémas

Si l'on souhaite explorer plus en avant la navigation visuelle, il est utile de mentionner l'architecte Kevin Lynch qui inventa le concept de « wayfinding »[58] (ou signalétique directionnelle) en 1960. Ce terme désigne un *« dispositif conjoint d'orientation, d'identification et d'informations visuelles dans un lieu »*, discipline qui obéit à trois critères :

- permettre à l'être humain de répondre précisément à la question : où suis-je ? ;
- lui donner les moyens de prendre une décision d'orientation ;
- lui fournir une image mentale de l'espace dans lequel il évolue.

Le « wayfinding » se concrétise dans la vie courante par le recours à la signalétique et aux pictogrammes. Rapporté au processus de consultation en ligne, (Morizio, 2006) fait une analogie avec un cheminement dans un labyrinthe : *« les premiers pas étant derrière soi, devenus invisibles, et la lente progression à travers des écrans successifs ne permettant pas d'apercevoir la lumière. Il faut à ce labyrinthe un fil d'Ariane, pour revenir au début si on le souhaite, aussi pour avoir des repères par rapport aux différentes étapes par lesquelles on est passé. Car après le dialogue de sourds, on court le risque du parcours en aveugle »*.

S'il s'avère nécessaire de devoir orienter l'étudiant vers un type de ressources particulier, nous pensons que l'utilisation d'un langage de pictogrammes est un bon moyen pour y

[58] DELHOMME S.-A., *Wayfinding : le bon sens loin de chez soi*, Courrier International, 21 avril 2008 http://www.courrierinternational.com/chronique/2008/04/21/wayfinding-le-bon-sens-loin-de-chez-soi

parvenir. Les pictogrammes sont des réalisations graphiques à fort degré d'iconicité: ils figurent les objets du réel par analogie, mais ils se distinguent de l'icône proprement dite par leur fonction de communication. (Cooper, 1995) souligne que, dans nos environnements informatiques, les interfaces utilisateurs sont majoritairement basées non pas sur des métaphores, mais sur des idiomes : *« des objets façonnés tels que les boîtes servant à fermer les fenêtres, les fenêtres redimensionnables, les dossiers imbriqués à l'infini ou encore le cliquer-déposer sont des opérations non-métaphoriques qui n'ont aucun parallèle dans le monde réel. Ils tirent leur force uniquement de leur apprentissage idiomatique facile »*. (Cartier, 1990) écrivait déjà il y a vingt ans : *« Nous vivons le passage de l'interopérabilité des réseaux à celles des cultures ; à l'avenir, les codes iconiques et la schématique joueront un rôle de plus en plus important parce que poussés par la mondialisation »*. Pour (Vaillant, 1999), le signe iconique laisse l'espoir d'être reconnu par n'importe qui grâce à un travail de raisonnement sur certains éléments intrinsèques du signifiant. Il est donc idéalement le candidat à ce rôle de vecteur de communication compréhensible universellement, sans apprentissage. Pour (Barrier, 2000), en contrôlant avec juste mesure certaines composantes telles que la symétrie (sur l'axe vertical), l'équilibre vectoriel, la répartition optimum du poids visuel, la binarité du contraste, le pictogramme s'impose au champ perceptif et laisse une trace durable en mémoire qui en facilitera la reconnaissance. Le rapport entre la figure et le fond doit être parfaitement équilibré afin de permettre une intégration visuelle correcte.

Nous donnons deux exemples de l'utilisation de pictogrammes pour l'aide à la navigation dans un contexte de recherche. Le premier est un moteur de recherche destiné aux enfants, *UpToTen*[59] (Figure 34). La recherche peut se faire de manière classique en renseignant le champ de recherche situé en haut de page ou bien en sélectionnant une icône correspondant à ce que l'on recherche. Le second exemple, *Symbaloo*[60], est un service de page d'accueil personnalisable à base d'icônes qui s'articule autour de blocs interactifs, parmi lesquels plusieurs moteurs de recherche (Figure 35).

[59] http://www.uptoten.com/
[60] http://www.symbaloo.com/

Figure 34 : UpToTen et son interface iconique

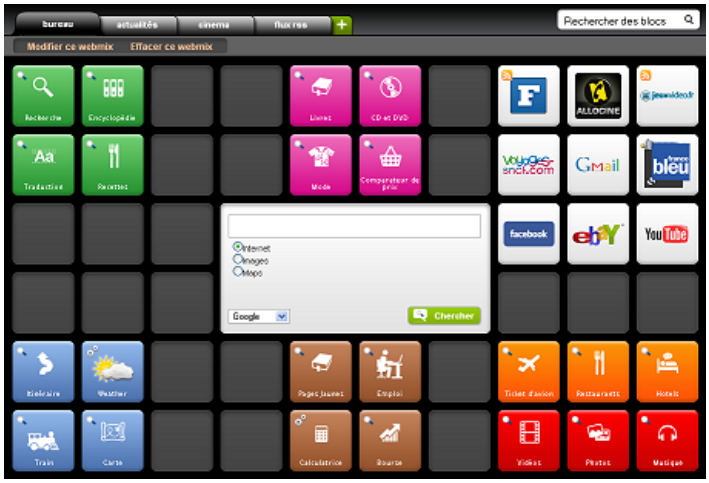

Figure 35 : Symbaloo

Une autre approche intéressante pour se repérer dans un espace est le recours à la schématique pour guider l'utilisateur. Nous donnerons deux exemples qui pourraient être envisagés pour la médiathèque « la Fusée » : le plan d'une fusée (Figure 36) et le plan d'un bâtiment (Figure 37). Une autre métaphore qui tend à être de plus en plus utilisée ces dernières années est celle du plan de métro[61], en particulier celle du métro de Londres qui a servi de modèle aux autres plans du même type. Une telle carte pourrait être conçue pour représenter visuellement un thésaurus. Par exemple, le plan de la Figure 38 illustre les rôles principaux des œuvres de Shakespeare.

[61] « Les Anglo-saxons ont surnommé cette représentation diagramme schématique. L'information géographique n'est pas précise, au contraire, lignes droites et angles fixes servent de base au dessin, quant aux distances, elles sont relatives. »
http://www.le-cartographe.net/index.php/blog/archives/107-la-representation-cartographique-du-metro

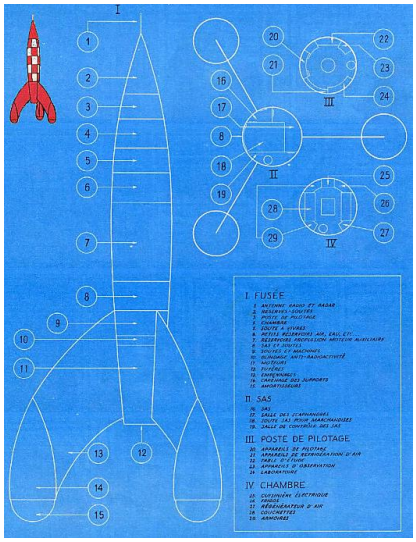

Figure 36 : Plan d'une fusée, tirée de la bande-dessinée « On a marché sur la Lune » (Hergé)

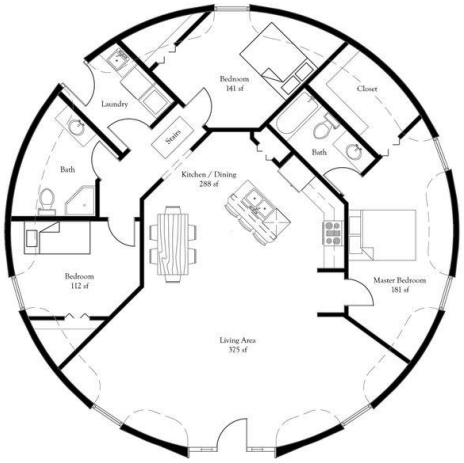

Figure 37 : Exemple de représentation d'un bâtiment selon une coupe horizontale

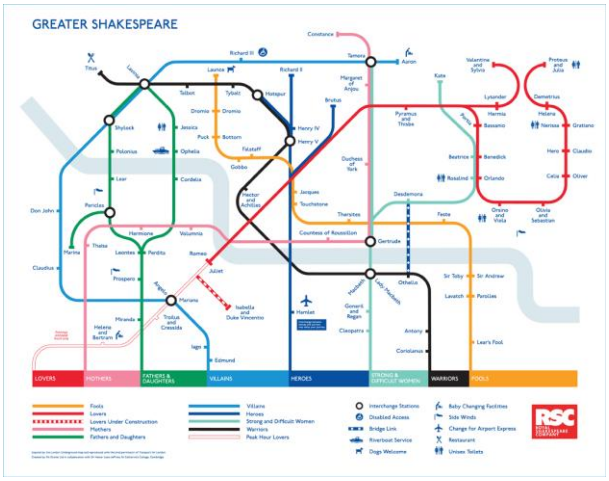

Figure 38 : *Greater Shakespeare Map*, par la Royal Shakespeare Company

Une solution intéressante qui mérite d'être étudiée dans le cadre de la médiathèque « la Fusée » est l'outil de recherche visuelle proposé par la *London Business School*. Baptisée *ConceptSpace*[62], l'application propose de naviguer dans une interface à base de sphères colorées qui utilise la métaphore des planètes, des satellites et des orbites. Grâce à un thésaurus spécialisé, l'interface graphique aide à localiser précisément le sujet de recherche, ce qui conduit à des résultats de recherche pertinents (Figure 39 et Figure 40).

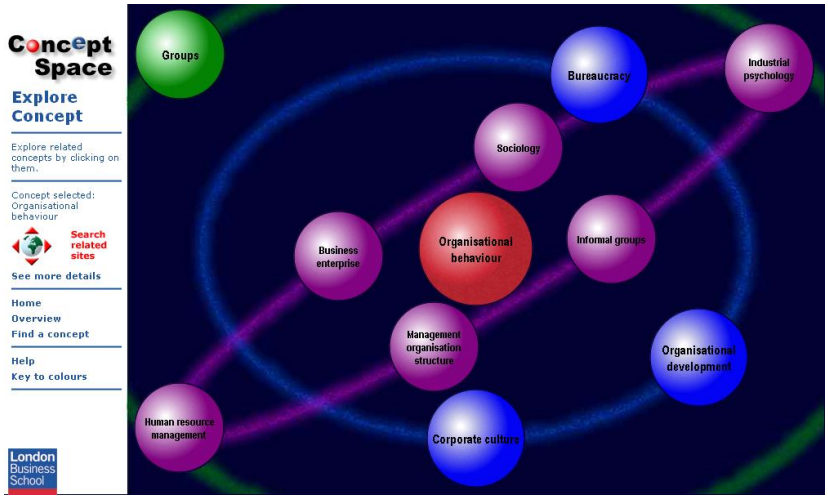

Figure 39 : ConceptSpace, un thésaurus visuel pour trouver des documents sur des sujets précis

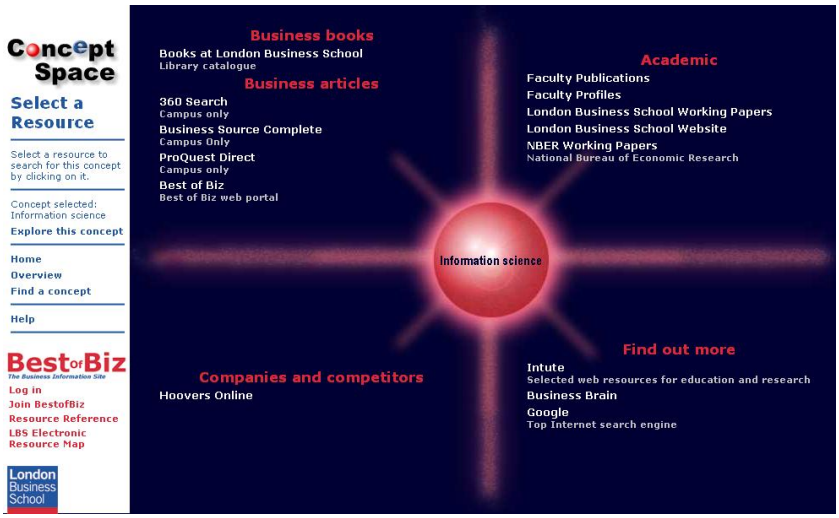

Figure 40 : ConceptSpace, page de sélection des bases sur un terme donné

[62] http://conceptspace.london.edu/

Le projet *Visual Catalog*[63], initialement conçu par Fabrice Papy et Sophie Chauvin, utilise des métaphores graphiques de représentation de la distribution des résultats de recherche. Implémenté depuis 2007 à l'Université d'Artois (Figure 41), l'interface permet la recherche dans les 16000 libellés de la classification Dewey et la « déambulation » virtuelle dans les rayonnages des deux salles de la bibliothèque universitaire. L'outil tire les leçons du manque d'accessibilité des OPACs classiques et du manque d'habilité instrumentale de l'usager. Composée de trois modules (recherche, exploration de la classification et localisation virtuelle), l'application fonctionne en complémentarité et permet aux usagers diverses formes d'accès. (Maisonneuve, 2008) souligne que l'analyse des transactions montre que 70% des requêtes aboutissent à des résultats. Les enquêtes par questionnaire ont montré que le *Visual Catalog* permettait aux usagers de trouver des ouvrages dont ils ne soupçonnaient pas l'existence (ibid).

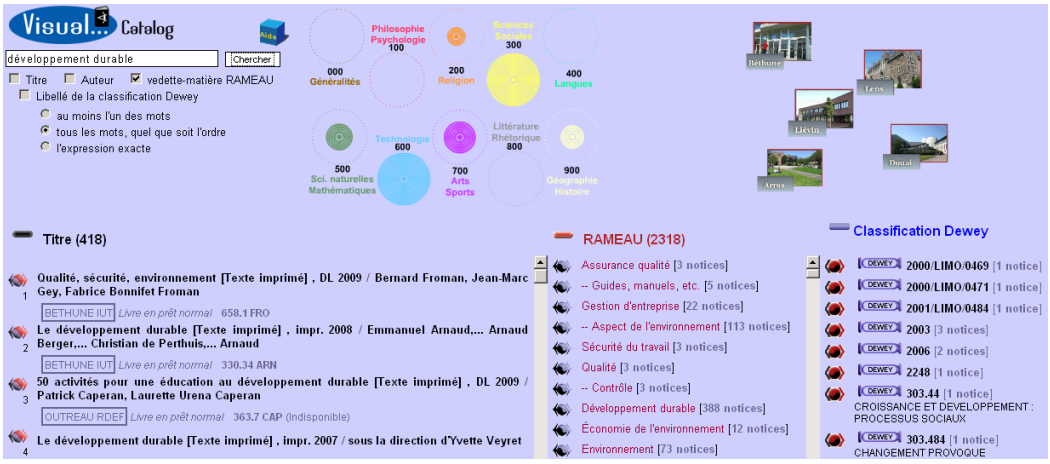

Figure 41 : L'interface du Visual Catalog tel que présenté sur le site de l'Université d'Artois

(Chevallet, 2002) recense plusieurs langages visuels pour spécifier des requêtes, notamment à l'aide de diagrammes de Venn[64] pour représenter les requêtes booléennes. Nous donnons ici en illustration l'application *VQuery* (Figure 42), de (Jones et McInnes, 1998). Les résultats de (Hertzum et Frokjaer, 1996) ont montré que les erreurs de syntaxe sont moindres et que cette méthode offre un support adapté au processus itératif de recherche.

[63] http://docnum.info/. On trouvera une fiche complète du produit dans *Le catalogue de la bibliothèque à l'heure du Web 2.0*, Tosca Consultants, ADBS éditions, 2008, pp. 241-259

[64] Schémas géométriques utilisés pour représenter des relations logico-mathématiques. http://fr.wikipedia.org/wiki/Diagramme_de_Venn

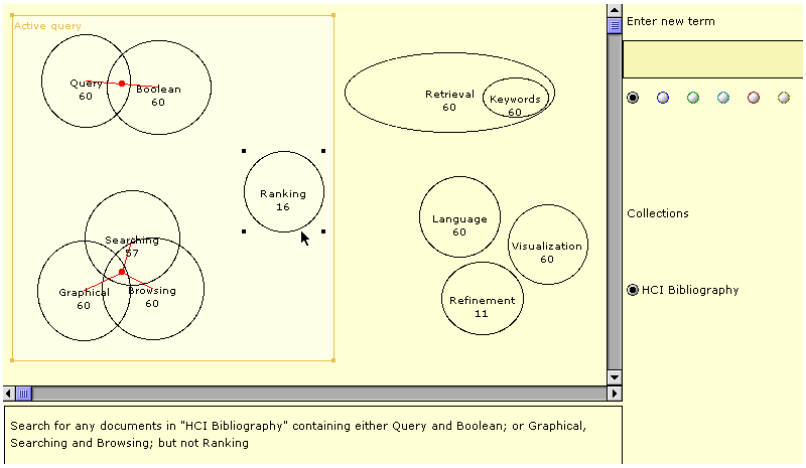

Figure 42 : VQuery (Jones and McInnes, 1998) interface à base de diagrammes de Venn pour la spécification de requêtes booléennes.

Selon (Bonnel, 2008), plus le nombre de documents retrouvés augmente, plus l'identification des informations pertinentes dans l'ensemble de ces documents peut être délicate, difficile et requérir un effort cognitif considérable. Pour l'auteur, on s'oriente vers l'émergence d'interfaces de visualisation multiples qui permettent de visualiser le même ensemble de résultats selon différents points de vue qui répondent aux besoins de plus de tâches ou de plus d'utilisateurs. Le projet MACE, cité plus haut, en est un parfait exemple.

En ce qui concerne la lecture des documents eux-mêmes, (Morizio, 2006) rappelle que la lecture du document électronique se prête plus facilement à un parcours erratique, car il s'agit souvent moins de lire pour trouver une information que l'on cherche, que pour découvrir. Du côté des solutions, on note l'apparition d'interfaces inspirées de la lecture papier. Par exemple, *Google Fast Flip*[65] permet d'accéder à des articles présentés sous la forme de pages web et de les feuilleter virtuellement. *Delicious BrowseBar*[66] offre la navigation contiguë à l'outil de partage de signets pour consulter les pages web mises en favoris, les unes après les autres. Le moteur de recherche *Widexplorer*[67] applique cette logique pour proposer, à partir d'une seule page web, l'accès à plusieurs moteurs de recherche placés horizontalement les uns à la suite des autres. Cependant l'implémentation d'une telle solution pour la consultation d'articles issus de revues électroniques semble compliquée du fait de l'usage du format PDF par défaut. Néanmoins, il s'agit d'un faux problème, car des solutions basées sur Flash sont

[65] http://fastflip.googlelabs.com/
[66] http://blog.delicious.com/blog/2010/01/filter-and-browse.html
[67] http://www.widexplorer.com/

déjà implémentés pour la presse et les magazines (à l'image du service *Issuu*[68], Figure 43) et les livres électroniques (par exemple *l'Harmathèque*[69]). Notons au passage qu'EBSCO a récemment mis en place un nouveau lecteur PDF intégré à sa plateforme *EBSCOHost*. L'outil est accompagné d'un navigateur contextuel qui permet d'accéder aux différents chapitres et autres liens qui figurent dans les documents (Figure 44). Dans le domaine, *Google Books* offre certainement l'interface la plus pratique pour la consultation d'ouvrages et de magazines (Figure 45) avec des fonctionnalités sociales : possibilité de choisir différents modes d'affichage (une page, double page, vignettes), zoom, consultation plein écran, accès à la table des matières, boutons de consultation (page suivante, page précédente), possibilité d'ajouter des commentaires et de noter le document, permalien[70] pour partager le document et possibilité d'ajouter le document à « ma bibliothèque ».

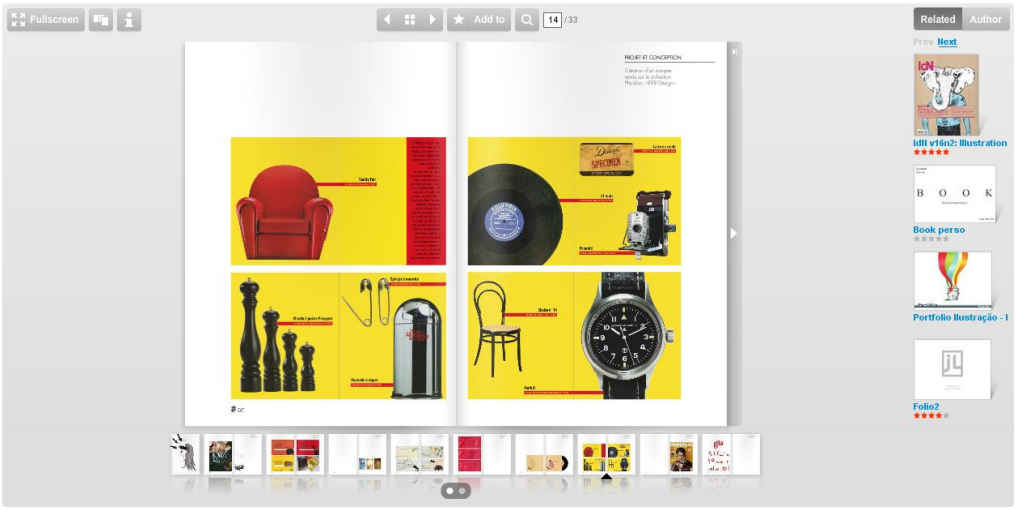

Figure 43 : Issuu et son interface de consultation de magazines

[68] http://issuu.com/
[69] http://www.harmatheque.com/
[70] De l'anglais Permalink, contraction de « Permanent Link ». Adresse Web (URL) conçu pour référer un élément d'information et pour rester inchangé de façon permanente, ou du moins, pour une certaine période.

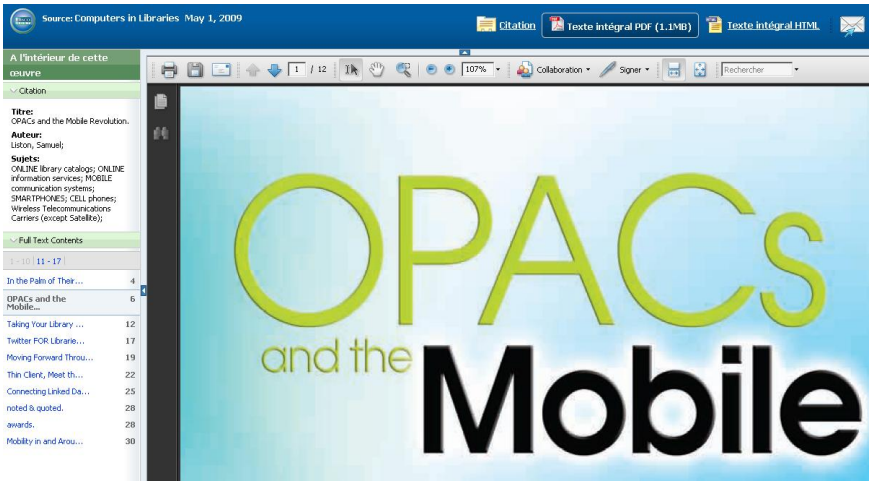

Figure 44 : Le nouveau lecteur PDF intégré à la plateforme EBSCOHost

Figure 45 : Interface de lecture de Google Books

3.5 L'utilisation de facettes pour l'accès aux ressources

La facette est une caractéristique, un trait, un indicateur ou un critère, servant à diviser une classe ou un ensemble d'objets en sous-ensembles homogènes (Hudon, 2007). A la différence des systèmes hiérarchiques qui attribuent aux sujets des séquences fixes, les systèmes de classification à facettes attribuent aux sujets les aspects, les propriétés ou les caractéristiques d'une classe ou d'un sujet spécifique, définis clairement, exclusifs les uns des autres et exhaustifs dans leur ensemble (Taylor, 2000). La classification à facettes offre

plusieurs chemins pour localiser un sujet. En utilisant différents aspects/facettes d'un sujet, les utilisateurs peuvent restreindre la recherche et localiser le sujet plus facilement, bien que chaque système hiérarchique et à facettes ait ses points forts et ses faiblesses (Kwasnik, 1999).

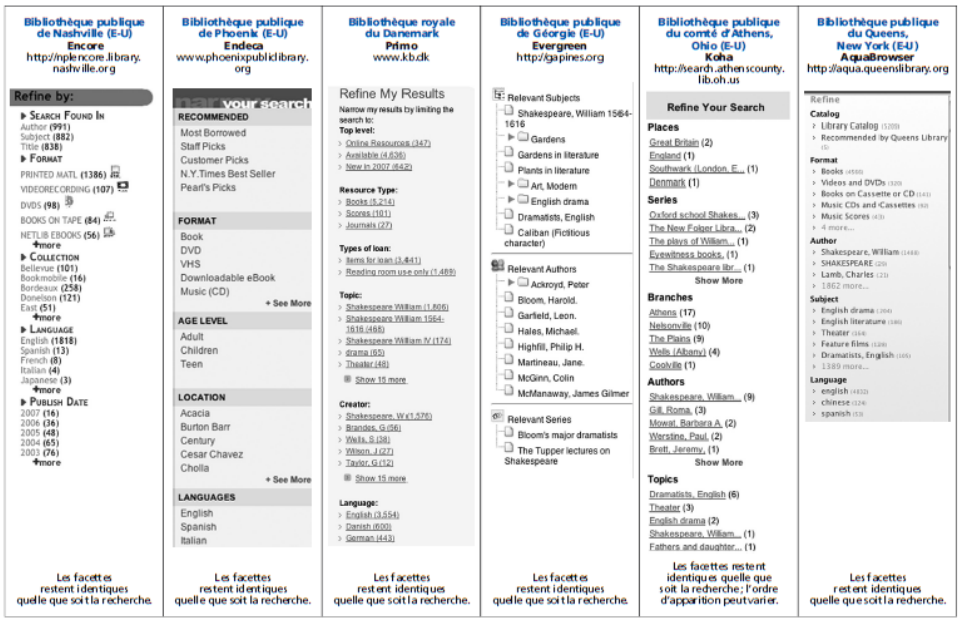

Figure 46 : Exemples de facettes proposées par différentes bibliothèques (Maisonneuve et Touitou, 2007)

La classification à facettes permet de présenter l'information d'une manière souple et structurée. Elle offre la possibilité de naviguer logiquement vers l'ensemble des ressources pertinentes pour répondre à un besoin ou à un intérêt lors de la recherche d'information (Figure 46). Selon (Hudon, 2007), les principaux avantages de cette solution sont :

- de permettre de représenter certains sujets beaucoup plus finement que les schémas énumératifs (tels la CDD[71] et la CDU[72]), à partir d'un nombre total de classes distinctes beaucoup moins grand ;
- d'être moins rigide, car elle s'adapte facilement au renouvellement des concepts : on peut modifier les isolats rattachés à une facette, ou même ajouter une facette, sans toucher à la structure globale du système ;

[71] Classification décimale Dewey <http://fr.wikipedia.org/wiki/Classification_d%C3%A9cimale_de_Dewey>
[72] Classification décimale universelle <http://fr.wikipedia.org/wiki/Classification_d%C3%A9cimale_universelle>

- de bien se prêter à l'automatisation de la recherche, puisqu'un sujet peut être repéré à partir de chacune de ses facettes et de la représentation de celle-ci.

La flexibilité de la structure à facettes est aussi liée à la façon dont la structure est utilisée non seulement pour l'organisation de ressources en groupes homogènes pertinents, mais également pour la description de chacune des ressources constituant la collection à classifier. Pour (Gavin, 2008), ce qui conditionne le succès d'une navigation à facettes, ce sont des données de qualité :

- données homogènes (catalogues cohérents) ;
- données riches (indexation matières, etc.) ;
- données codées (format MARC, codes de langues, etc.) ;
- données normalisées (auteurs sous forme normalisée, etc.) ;
- données compatibles (selon normes internationales) ;
- identificateurs (ISBN, DOI, etc.)

Pour l'utilisateur, de par son aspect multidimensionnel, la classification à facettes peut néanmoins apporter quelques difficultés comme le souligne (Balmisse, 2007) :

- elle peut être la cause de désorientations pour les utilisateurs ;
- la navigation n'est pas scénarisée. Il n'y a pas de dimensions plus importantes que les autres, ni d'ordre préétabli entre les facettes. C'est à l'utilisateur qu'incombe la lourde tâche de trouver le chemin adéquat vers l'information pertinente ;
- les facettes doivent être choisies les unes après les autres ;
- la conception d'un système à facettes performant peut s'avérer longue et délicate.

Le moteur actuel de la médiathèque du campus SKEMA de Lille fonctionne avec la solution *AquaBrowser* (Figure 47 et Figure 48), créée par la société Medialab (rachetée par SerialsSolutions / ProQuest en 2009). Lancé en 1997, *AquaBrowser* a su s'adapter pour proposer une interface moderne et conviviale. Calquée sur le modèle des moteurs de recherche du Web, celle-ci propose une boîte de recherche unique, une réponse quasi instantanée, une gestion automatique des fautes de frappe et la visualisation des cooccurrences. Les résultats de chaque facette sont triés par ordre d'importance et indiquent le

nombre d'occurrences des documents concernés. En plus de l'affinage des résultats, les facettes permettent une découverte du fonds grâce à une vision différente et pertinente des principaux aspects liés à la recherche originale. En juillet 2009, *AquaBrowser* comptait environ 700 bibliothèques parmi ces clients[73].

Figure 47 : Page de résultats d'AquaBrowser sur le portail de la médiathèque de Skema BS / ESC Lille

[73] AquaBrowser Becomes Top Discovery Layer for Libraries
http://www.proquest.com/en-US/aboutus/pressroom/09/20090721.shtml

Figure 48 - AquaBrowser : Notice d'un ouvrage

Aurélie Duclos, animatrice de communauté au sein de la médiathèque du campus de Lille, a présenté en janvier 2010 un retour d'expérience sur l'usage d'*AquaBrowser* au sein de SKEMA[74]. Il en ressort que :

- Certains étudiants comprennent plus ou moins comment fonctionne le nuage de mots-clés; d'autres pas du tout, mais ils trouvent l'outil suffisamment intéressant pour chercher à s'en servir.
- En ce qui concerne les facettes, certains critères sont plus utilisés que d'autres (ex : disponibilité, format, sujet, année). Cela équivaut à l'usage d'opérateurs booléens, mais sans nécessité d'apprendre une quelconque syntaxe.
- Les résultats de recherche produisent très peu de silence.
- L'affichage des couvertures extraites du site Amazon est très appréciée, mais il y a parfois une confusion dans les zones à sélectionner.
- Le tri par pertinence proposé par défaut est très bon, il y a très peu de changements de critère de tri.

A noter que l'interface d'administration de l'outil permet d'obtenir d'autres informations sur les pratiques des utilisateurs.

[74] http://www.slideshare.net/Auremims/aquabrowser-fonctionnalits-mise-en-place-apports-aux-usagers/

4 LES BIBLIOTHEQUES A L'HEURE DU NUMERIQUE

4.1 Vers une bibliothèque « 2.0 »

En raison de la multiplicité des supports d'information, les bibliothèques se transforment peu à peu en médiathèques. Les demandes portent de plus en plus sur le multimédia, mais aussi sur l'information spécialisée, notamment les périodiques en ligne. Les fonds électroniques, les postes de lecture assistée par ordinateur, la délocalisation des sources d'information, l'accès aux documents à distance, la numérisation des collections, autant d'éléments qui font évoluer les bibliothèques en profondeur, les faisant passer du matériel à l'immatériel, du réel vers le virtuel et de l'analogique au numérique (Mkadmi et al, 2008). C'est ainsi que les termes de bibliothèque numérique, bibliothèque virtuelle et bibliothèque hybride sont apparus. Cette mutation implique de nouveaux outils : des logiciels d'interrogation, de navigation et de recherche, des formats de représentation des documents qui facilitent l'échange des documents électroniques. Parmi les définitions que nous pouvons trouver sur les bibliothèques numériques, nous retenons celle-ci :

> *« Une bibliothèque numérique est un assemblage d'outils de numérisation, de stockage et de communication, de contenus et de logiciels nécessaires pour reproduire, émuler et étendre les services offerts par les bibliothèques traditionnelles fondées sur le papier et autres moyens de collecte, de catalogage, de recherche et de diffusion de l'information. Un système de bibliothèque numérique complet doit fournir les services essentiels d'une bibliothèque classique et aussi exploiter les avantages bien connus du stockage, de la recherche de la communication numériques. »* (Gladney et al, 1994)

On parle de bibliothèque virtuelle lorsqu'une bibliothèque permet à ses usagers l'accès à des ressources distantes. La bibliothèque numérique intègre les médias ou ressources numériques dont dispose la bibliothèque. En simplifiant, on peut dire que la bibliothèque hybride associe les trois systèmes : la bibliothèque traditionnelle, la bibliothèque numérique et

la bibliothèque virtuelle (Le Bescond, 2005). Les bibliothèques numériques en tant que dépôt de la connaissance peuvent être considérées comme des systèmes d'information complexes, dont le développement et l'utilisation requièrent de nombreuses solutions scientifiques, technologiques, méthodologiques, économiques, juridiques et autres. (Kalinichenko et al, 2002)

Selon Michael Casey[75], *« la bibliothèque 2.0 est un modèle qui reflète une transition interne au monde des bibliothèques dans la façon de fournir du service à leurs utilisateurs. Ce changement de direction sera spécialement manifeste dans l'offre électronique telle que la configuration de l'OPAC, les services en ligne de la bibliothèque, ainsi qu'un flux grandissant d'information en provenance de l'utilisateur ».*

4.2 Les portails documentaires

Le portail de bibliothèque est la somme des instruments développés par les bibliothèques sur lesquels on superpose des caractéristiques empruntés aux portails Internet (par exemple Yahoo) et aux portails d'entreprise (Le Bescond, 2005). 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 déchargement. 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 (Maisonneuve, 2003).

Pour (Arsenault et Paré, 2005), les principales fonctionnalités qui caractérisent le portail de bibliothèque se situent à quatre niveaux :

[75] Michael Casey, LibraryCrunch : Service for the Next Generation Library : A Library 2.0 Perspective <http://www.librarycrunch.com/>, cité par Muriel Amar, Véronique Mesguich, in *Le Web 2.0 en bibliothèques, Quels services ? quels usages ?*, Cercle De La Librairie, ISBN 2765409765

- la méta-recherche, qui permet de chercher simultanément dans plusieurs bases de données documentaires ;
- la personnalisation par la bibliothèque : contrôle de l'apparence du portail, ainsi que certains paramètres de recherche, et la possibilité de définir à l'avance des groupes de ressources destinées à être interrogées simultanément ;
- la personnalisation par l'usager : par exemple la possibilité de conserver l'historique de leurs recherches ou de créer des alertes ;
- l'authentification pour restreindre l'accès aux bases de données à certains usages ou groupes d'usagers.

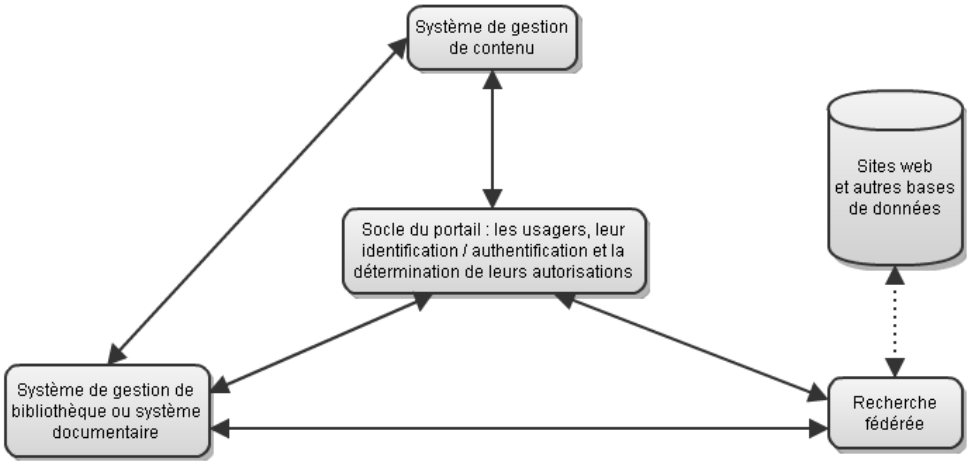

Figure 49 : L'articulation des principales briques d'un portail pour bibliothèque (d'après Maisonneuve, 2007)

Pour Dominique Lahary[76], *« le SIGB n'est plus le centre, c'est une composante de la constellation convergeant vers le portail de bibliothèque. D'où la recherche fédérée vers des données diverses, les services question-réponse, etc. Le tournant du siècle a vu l'émergence de nouveaux acteurs sur ces fonctions de portail, vers lequel nombre de fournisseurs de SIGB ont aujourd'hui élargi leur offre ».*

[76] Texier B. « le SIGB n'est désormais plus qu'une composante », entretien avec le bibliothécaire Dominique Lahary, Archimag n°210, décembre 2007
http://beta.archimag.com/article/%C2%AB-le-sigb-n%E2%80%99est-d%C3%A9sormais-plus-qu%E2%80%99une-composante-%C2%BB

4.3 Les OPACs de nouvelle génération

Apparus à la fin des années 70, les premiers OPACs[77] (Online public access catalog) étaient proposés en mode texte et consultables uniquement au sein de la bibliothèque. Désormais de plus en plus d'OPAC sont consultables en ligne, mais ceux-ci n'ont pas su évoluer pour s'adapter aux nouvelles pratiques des jeunes. A l'occasion d'une conférence donnée par Marc Maisonneuve sur les OPACS de nouvelle génération, la bibliothécaire Nadine Pestourie résumait sur son blog[78] au sujet des OPACs « classiques » : *« Ils ont un taux de silence redoutable, ne donnent rien à voir, n'incitent pas à la découverte, ne sont pas sexy, offrent peu d'aide au choix, ne sont pas encourageants pour l'usager qui doit se débattre avec un jargon inconnu »*.

Les OPACs de nouvelle génération ont pour objectif d'améliorer le processus de recherche, d'offrir une information aussi riche que possible et d'offrir aux lecteurs des services pour exploiter au mieux l'information trouvée. Comme l'explique (Maisonneuve, 2008) dans son ouvrage consacré au sujet, *« le principe fondateur de ce nouveau catalogue est d'offrir une recherche similaire à ce que propose Google. Outre l'adoption de l'ergonomie du moteur de recherche, avec généralement une fenêtre de recherche placée en haut d'un écran dépouillé, l'OPAC adopte également un principe de Google : peu importe le bruit, il ne faut pas qu'il y ait de silence. L'usager sait s'adapter au bruit, mais pas au silence »*. Alors que dans les OPACs actuels l'utilisateur est sollicité avant de commencer sa recherche, les nouveaux OPACs proposent une interface minimaliste et déplace l'étape de filtrage en amont (Figure 50). Maisonneuve (ibid) identifie six principes qui définissent ces nouveaux OPACs :

- l'écran de saisie de la requête évoque Google ;
- l'utilisateur doit pouvoir choisir son mode de prise de connaissance des résultats ;
- l'utilisateur s'accommode bien du bruit, mais mal du silence ;
- l'utilisateur doit pouvoir continuer sa recherche sans avoir à ressaisir la moindre information ;
- l'utilisateur doit voir le produit pour avoir envie de l'acheter ;

[77] Le terme français CIEL, pour *Catalogue informatisé en ligne*, a été proposé, mais n'est pas utilisé en pratique par les bibliothécaires.
[78] *OPACs de nouvelle génération*, billet publié le lundi 6 avril 2009, sur le blog de Nadine Pestourie http://lebibwebzine.blogspot.com/2009/04/opacs-de-nouvelle-generation.html

- les avis des autres usagers sont rassurants, il faut donc les mettre en avant.

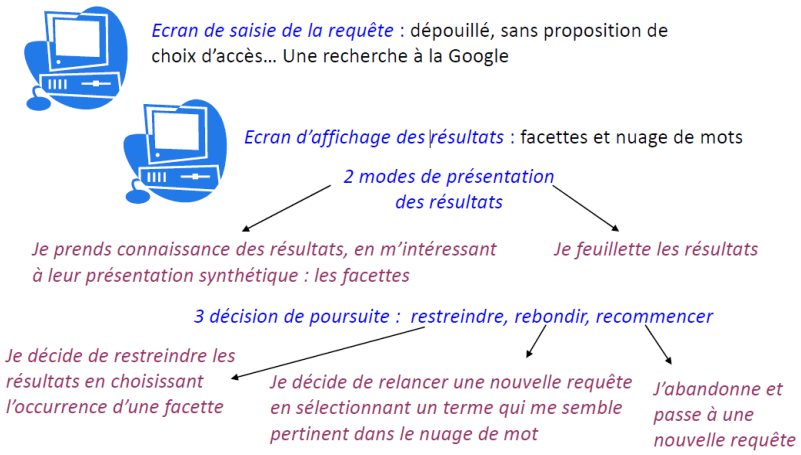

Figure 50 : L'utilisation d'un OPAC de nouvelle génération (Maisonneuve, 2009)

Certains OPACs proposent également des fonctionnalités inspirées des sites commerciaux, tels que le système de suggestion (« Les lecteurs qui ont emprunté ce livre ont aussi emprunté les livres suivants… »).

Ces OPACs sont constitués de composants qui interviennent dans la préparation des données faisant l'objet de recherches :

- le moteur d'indexation et de recherche ;
- le client OAI-PMH de l'OPAC (voir chapitre 5.2) ;
- les procédures automatisant l'import des données ;
- les chargeurs assurant la mise en correspondance des informations importées et de la base interne ;

puis ceux qui interviennent dans la recherche et l'affichage des informations :

- le socle de l'OPAC (dont l'interface utilisateur);
- le résolveur de liens[79] ;
- le module d'indexation et de recherche ;

[79] Il s'agit d'un logiciel qui convertit une adresse logique, un identifiant (par exemple un numéro ISBN) ou des métadonnées, en l'adresse URL physique du document recherché.

- le client Z39.50, SRU ou SRW ;
- le connecteur LDAP de l'OPAC.

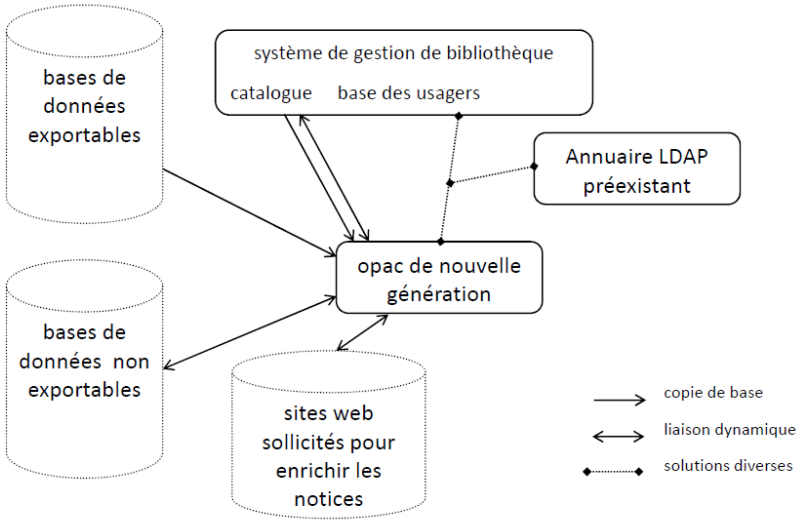

Figure 51 : L'architecture d'un OPAC de nouvelle génération (Maisonneuve, 2009)

Les nouveaux OPACs qui sont commercialisés avec des fonctionnalités améliorées sont dénommés dans les pays anglosaxons « discovery systems »[80]. En mars 2008, la *Digital Library Federation* (DLF) a ainsi réuni plusieurs gros acteurs du domaine pour bâtir une proposition de recommandation afin de favoriser le développement d'OPACs de nouvelle génération de plus en plus indépendants des systèmes de gestion de bibliothèque et l'enrichissement des services proposés par ces OPACs.[81]

[80] On trouve également les dénominations d' « integrated discovery system », « discovery services », « discovery interfaces », « discovery layer », « discovery layer interfaces », « discovery tools » et « resource discovery service ».
[81] Les recommandations qui sont issues de cette collaboration sont disponibles en ligne (décembre 2008). http://www.diglib.org/architectures/ilsdi/DLF_ILS_Discovery_1.1.pdf

5 LES SOLUTIONS DE RECHERCHE GLOBALE

5.1 La recherche fédérée

La recherche fédérée ou méta-recherche[82] (Figure 52) est un processus qui permet essentiellement d'interroger, de façon simultanée et par le biais d'une interface unique, une multitude de ressources hétérogènes telles que des catalogues, des bases de données, des dépôts institutionnels, des services agrégatifs de périodiques, des encyclopédies et d'autres sources d'information en format numérique (Arsenault et Paré, 2005). Pour l'utilisateur, l'avantage est qu'il n'a plus à se familiariser avec les interfaces natives des différents systèmes. C'est l'une des deux techniques utilisées pour consulter des bases réparties, l'autre étant l'architecture des entrepôts OAI. Généralement, l'outil de recherche fédérée demeure imparfait en la matière. Il ne dispose que d'un descriptif relativement sommaire des ressources électroniques auxquelles il donne accès et traite ces contrôles dans une logique de tout ou rien, c'est-à-dire sans prendre pleinement en compte les particularités de la ressource électronique (Maisonneuve, 2007). Le méta-moteur élimine habituellement les doublons et les présente dans une liste consolidée tout en offrant différentes possibilités de traitement des résultats (filtres, paniers, envoi par courriel, exportation, création de profils de recherche personnalisés et regroupement) (Robert, 2009).

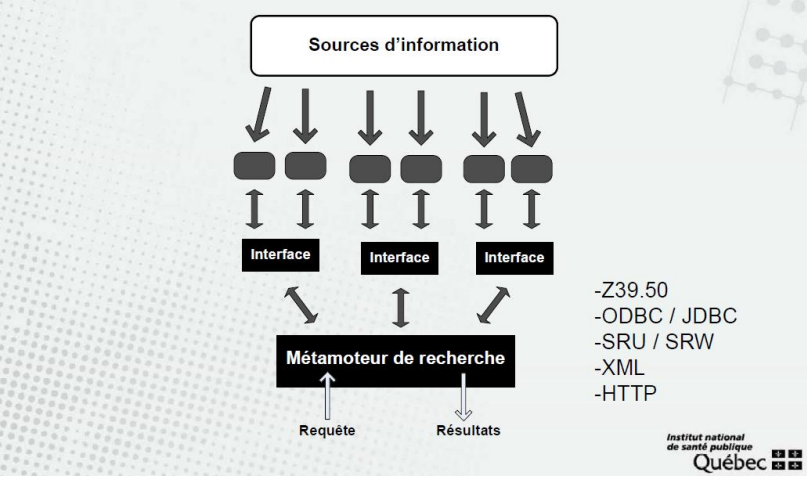

Figure 52 : Schéma représentant le principe de la recherche fédérée (Robert, 2009)

[82] Le terme méta-recherche est équivalent aux expressions recherche fédérée, recherche diffusée, recherche parallèle, recherche intégrée et recherche multi-bases ou multi-cibles. En anglais, on retrouvera notamment les expressions metasearch, federated search et broadcast search. (Arsenault et Paré, 2005)

Ce type d'outil est apparu en 1999 alors que le nombre de bases de données ne cessait d'augmenter. Le recours à la recherche fédérée dépend avant tout de la quantité de documents numériques accessible depuis des sources extérieures et des données locales. Le tableau présenté ci-après (Tableau 1) donne une idée de l'intérêt que cette solution propose en fonction du volume de données. Dans le cadre de SKEMA, les abonnements aux bases de données (plus d'une dizaine) représentent un volume important de documents accessible sous forme numérique. A titre d'exemple, la base *Factiva* propose, à elle seule, plus de 215 revues.

		Le catalogue et les données locales représentent…		
		20% du total des informations auxquelles l'OPAC donne accès (et les bases en recherche fédérée, soit le reste, 80%)	50% du total des informations auxquelles l'OPAC donne accès	80% du total des informations auxquelles l'OPAC donne accès
Les traitements documentaires (catalogage, indexation) pratiqués respectivement par la bibliothèque et les éditeurs des bases sollicitées en recherche fédéré sont…	**proches**	Niveau d'intérêt faible	Niveau d'intérêt moyen	Niveau d'intérêt élevé
	Largement différents	Niveau d'intérêt faible	Niveau d'intérêt faible	Niveau d'intérêt moyen

Tableau 1 : L'intérêt de recourir à la recherche fédérée en complément de l'exploitation des données locales (d'après Maisonneuve, 2008)

D'un point de vue technique, les échanges entre le module de recherche fédérée et les bases de données sont pris en charge par des connecteurs adaptés aux caractéristiques des bases d'information sollicitées en s'appuyant sur divers protocoles de recherche et de récupération d'information (Z39.50, SRU, SRW, etc.), sur les langages de requête qui y sont associés (Type-1 Query pour Z39.50, CQL pour SRU et SRW, etc.) et sur une structure normalisée des données récupérées (le plus souvent des formats Marc pour Z39.50, différents

formats de données bibliographiques exprimés en XML, notamment MarcXML ou des métadonnées Dublin Core dans le cas de SRU/SRW…) (Maisonneuve, 2007).

Le problème qui est apparu rapidement avec la recherche fédérée, c'est qu'avec l'augmentation de documents dans les collections, les recherches se sont vues ralentir, ne parvenant plus à retourner les résultats de manière quasi-instantanée. C'est l'une des raisons qui a freiné l'implémentation massive de ce type d'outil. Un des autres points qui pose problème est l'homogénéisation et le classement des résultats. Michael Dunne, responsable du département numérique de la bibliothèque de l'université de Lancaster écrivait récemment[83] : *« Notre bibliothèque a implémenté MetaLib[84] en 2005. A l'époque, c'était la solution la plus à la pointe en matière de recherche fédérée pour les bibliothèques. Cinq ans plus tard, la solution semble démodée, bien que fortement utilisée, il faut bien reconnaître qu'elle n'est pas très populaire auprès de nos utilisateurs (y compris les bibliothécaires). Les utilisateurs préfèrent utiliser Google. Ils ne se servent de MetaLib que lorsqu'ils savent que quelque chose de spécifique n'est disponible qu'à travers l'application, ou quand ils ont été dirigés par des tuteurs. L'exécution des requêtes est lente, parfois les recherches n'aboutissent pas, le classement ne saute pas aux yeux, et de par la complexité, les utilisateurs échouent à explorer les alternatives proposées par MetaLib. La majorité utilise la recherche rapide composée des bases Academic Search Complete, JSTOR, et Web of Science ».*

La recherche fédérée a également pour inconvénient de ne pas permettre la recherche multicritères sur les bases sources. Dans le même ordre d'idée, la recherche sur le champ « sujet » peut par exemple poser des problèmes à partir du moment où les champs « sujet » des différentes bases ne sont pas gérés de la même manière. Etienne Cavalie[85], bibliothécaire français, note de son côté *« Je crois que, au regard de l'investissement financier nécessaire, une recherche fédérée où l'utilisateur doit d'abord chercher les sources à sélectionner n'est pas rentable. La démarche de retrouver les ressources intéressantes nécessite de comprendre un outil souvent jugé trop complexe, et incite l'internaute à interroger directement les ressources qu'il connaît. »*, et de rajouter *« Contraindre l'utilisateur à choisir les bases qu'il va interroger, c'est lui imposer de comprendre ce qu'il est en train de faire avant de le faire*

[83] *Discovery Tools – Options for the future*, blog de Michael Dunne, billet publié le 12 février 2010 http://mdunne.wordpress.com/2010/02/12/discovery-tools-%E2%80%93-options-for-the-future/
[84] http://www.exlibrisgroup.com/category/MetaLibOverview
[85] CAVALIE E., *Pour ou contre la recherche fédérée ?*, billet sur le blog Bibliothèques [reloaded], 8 novembre 2008 http://bibliotheques.wordpress.com/2008/11/08/pour-ou-contre-la-recherche-federee/

(et de maîtriser un outil — un de plus — complexe). C'est donc la meilleure manière de l'inciter à aller voir ailleurs ».

5.2 La recherche unifiée

Les solutions de recherche unifiée[86] sont apparues très récemment afin de répondre aux problèmes rencontrés avec les solutions de recherche fédérée, où à chaque requête de l'utilisateur l'application doit aller interroger les différentes bases une à une. Avec ce nouveau type d'outil, les documents issus des différentes bases sont indexés dans un seul silo (on parle également de recherche unifiée et de recherche verticale). Les quelques sociétés qui commencent à proposer une solution de ce type passent des accords au cas par cas avec les éditeurs, afin de collecter les métadonnées auprès de ceux-ci. Il est également possible d'ajouter à cet index unique des contenus tiers, comme le catalogue de la bibliothèque local ou d'autres types de contenus (thèses, mémoires, etc.).

Ces fonctionnalités sont rendues possibles grâce à l'utilisation de normes d'interopérabilité qui permettent au système d'échanger des ressources avec d'autres systèmes grâce à l'utilisation de langages et de protocoles communs. Le protocole OAI-PMH (Open Archives Initiative's Protocol for Metadata Harvesting), créé en 1999, permet ainsi les échanges de données entre les fournisseurs de données et les fournisseurs de services. Les fournisseurs de données déposent leurs métadonnées sur un serveur web appelé « entrepôt » et les fournisseurs de service collectent (ou « moissonnent ») ces données pour les intégrer à l'index de leurs propres bibliothèques numériques. La Figure 53 illustre ainsi le mécanisme mis en œuvre pour un moteur de recherche unifiée.

[86] On parle là aussi de solutions de « discovery », parfois d' « unified discovery ». Il n'existe pas à l'heure actuelle de traduction française au terme qui correspond à ce type d'outil. Une traduction possible serait « outil de signalement et de recherche global ».

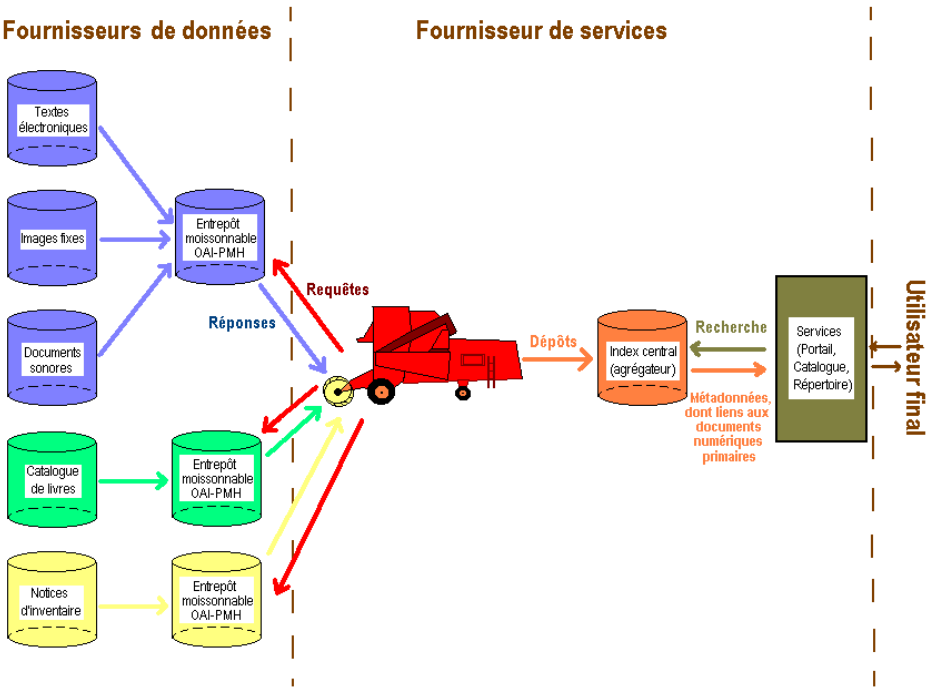

Figure 53 : Architecture fonctionnelle de services OAI pour la recherche unifiée[87]

Afin de faire le bon choix dans l'acquisition d'une solution de recherche unifiée[88], il est nécessaire de prendre connaissance des différentes fonctionnalités qui sont proposées et d'identifier lesquelles sont essentielles. Les principales fonctionnalités que l'on va retrouver dans ce type d'outil sont :

- un seul point d'entrée pour interroger les différents contenus ;
- l'application de moissonnage (qui traite les requêtes OAI-PMH) ;
- la solution d'authentification qui doit être transparente pour l'utilisateur[89] ;
- le support des différents standards (MARC, UniMARC, DublinCore, Z39.50, …) ;
- la détection des doublons ;
- le système de classement des résultats (algorithme de tri plus ou moins efficace) ;

[87] Source : http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm

[88] Andy Ekins (Christ Church University, Royaume-Uni) et Lukas Koster (Library of the University of Amsterdam, Pays-Bas) ont créé un site qui recense et compare les solutions de « discovery » du marché. http://sites.google.com/site/urd2comparison/

[89] L'authentification unique (ou identification unique ; en anglais Single Sign-On ou SSO) est une méthode permettant à un utilisateur de ne procéder qu'à une seule authentification pour accéder à plusieurs applications informatiques (ou sites Web sécurisés). http://fr.wikipedia.org/wiki/Sso

- la normalisation des résultats (pour rendre ceux-ci cohérents, par exemple sur la manière de traiter les caractères accentués) ;
- la classification à facettes ;
- la gestion de l'indexation personnelle (folksonomy[90]) ;
- un nuage de mots (« tag cloud ») ;
- la personnalisation de l'interface utilisateur ;
- la recommandation (système de notation et d'ajout de commentaires) ;
- la gestion du multilinguisme ;
- la possibilité de gérer les consortiums[91] de bibliothèques ;

Le premier éditeur à avoir mis sur le marché une solution de recherche unifiée est l'américain SerialSolutions (filiale de ProQuest) avec *Summon*[92]. Le produit a d'abord été testé auprès des bibliothèques américaines de *Dartmouth College* et *Oklahoma State University*, avant d'être proposé commercialement en juillet 2009. Aujourd'hui une vingtaine d'universités ont fait le choix de cet outil, parmi lesquelles *Grand Valley* et *Western Michigan University* (États-Unis), Sydney (Australie), Calgary (Canada) et *Eramus University* (Pays-Bas). En France, c'est l'université Paris 7 qui est en train de faire les premiers essais.

Les premiers retours d'expérience montrent que *Summon* a un réel impact sur l'accès aux ressources numériques des bibliothèques. L'université de Grand Valley a ainsi constaté une importante augmentation du taux de rebond vers les bases des journaux électroniques (Figure 54). Le produit fait l'objet d'améliorations régulières, afin de répondre à certains problèmes rencontrés. C'est le cas par exemple en ce qui concerne la rapidité d'intégration des nouveaux contenus. D'autres problèmes ne sont pas le fait de l'outil lui-même, mais sont liés aux fournisseurs de données : structure et manque de richesse des métadonnées, absence de lien direct aux articles, … Un autre problème potentiel porte sur l'accès depuis un terminal mobile. Il faut dans ce cas que l'éditeur propose les articles dans un format de consultation prévu à cet effet. Dans le cas contraire, *Summon* bénéficie d'une fonctionnalité qui permet de

[90] Néologisme désignant un système de classification collaborative décentralisée spontanée, basé sur une indexation effectuée par des non-spécialistes.
http://fr.wikipedia.org/wiki/Folksonomie
[91] Groupement d'entreprises juridiquement indépendantes réunissant, sous une direction unique, des moyens financiers ou matériels en vue d'exécuter en commun des opérations déterminées.
[92] http://www.serialssolutions.com/summon/

se constituer un « panier » des articles que l'on souhaite consulter. L'application peut ensuite envoyer un mail à l'utilisateur qui pourra les consulter ultérieurement depuis un poste fixe.

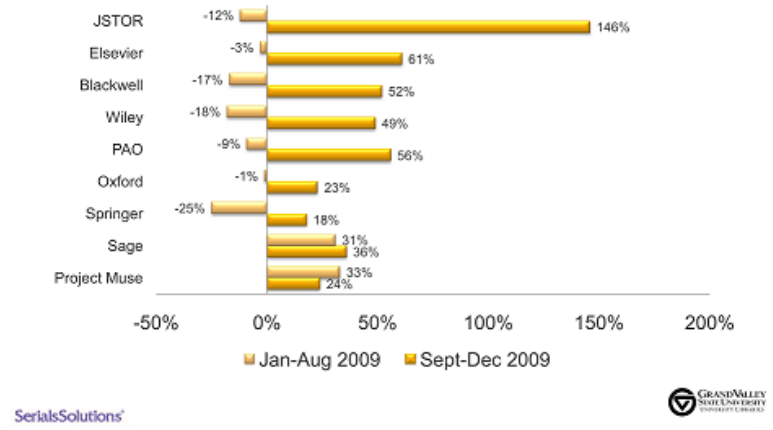

Figure 54 : Statistiques d'usage des bases à Grand Valley avant et après le passage à Summon

Techniquement parlant, *Summon* est hébergé sur les serveurs de SerialsSolutions, aux États-Unis. De fait, l'interface de recherche (en particulier la page des résultats) se trouve non pas sur le site de la bibliothèque, mais sur le site de *Summon* (exemple : http://dartmouth.summon.serialssolutions.com/). Pour autant, la bibliothèque est toujours en mesure de proposer le champ de recherche depuis son site ou depuis un « widget ». C'est seulement une fois que l'utilisateur a lancé la recherche qu'il est redirigé vers le site de *Summon* (Figure 55). *AquaBrowser* fonctionne de la même manière et il est à noter que les deux outils peuvent être associés pour bénéficier des avantages de chacun (ce qui permettrait de conserver l'interface actuelle dans le cas de « la Fusée »). *Summon* propose également un accès à son moteur depuis une API[93], ce qui permet aux bibliothèques qui le souhaitent de fournir à leurs usagers une interface qui leur est propre.

En ce qui concerne les métadonnées, *Summon* ne conserve qu'un seul enregistrement par document, pour éviter les duplications dans les résultats de recherche. Lorsque plusieurs versions de métadonnées sont reçues (de différentes bases), les enregistrements sont rassemblés pour ne former qu'une seule entrée.

[93] Il s'agit d'une interface de programmation applicative, un ensemble de commandes externes publiées par un éditeur et permettant de recourir aux fonctions d'un logiciel depuis un autre logiciel.
http://fr.wikipedia.org/wiki/Interface_de_programmation

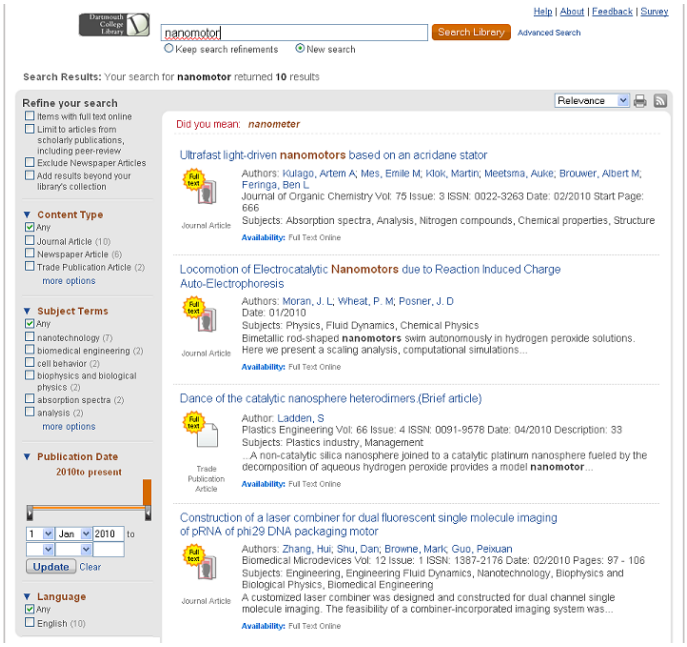

Figure 55 : Page de résultats de Summon

Parmi les autres solutions de recherche unifiée du marché qui voient le jour, nous pouvons citer *Primo*[94] de la société Ex-Libris. C'est l'offre retenue par l'université de Versailles (en cours d'implémentation), *EBSCO Discovery Service*[95] d'EBSCO et *Worldcat Local*[96] de l'OCLC. Dans le cas de *Primo* et de *Worldcat Local*, il s'agit semble-t-il d'une évolution de l'offre de recherche fédérée vers la recherche unifiée.

Il est plus que probable que le moissonnage des métadonnées des éditeurs de bases de données se généralise dans l'avenir au détriment de la recherche fédérée. Néanmoins, il reste encore de nombreuses bases qui ne proposent de moissonnage, c'est par exemple le cas de *Factiva* qui indexe le contenu de la presse. Dans ce cas précis, lorsque le contenu ne peut être indexé dans le silo unique, les moteurs ont recours à la recherche fédérée. D'autre part, les prestataires du marché étant tous américains, ils commencent à peine à signer des accords avec les éditeurs français. Autant de raisons qui font dire qu'il est préférable d'attendre encore quelques temps avant de choisir une solution de recherche unifiée.

[94] http://www.exlibrisgroup.com/category/PrimoOverview
[95] http://www.ebscohost.com/discovery
[96] http://www.oclc.org/worldcatlocal/

6 CONCLUSION ET PERSPECTIVES

Bien que l'offre ait littéralement explosé ces dernières années, (Ihadjadene, 1999) et (Spink, 2001) ont montré que les ressources sont sous-utilisées dans les catalogues en ligne. Pour avoir moi-même était confronté durant mon stage à la recherche d'information auprès de différentes sources, j'ai pu constater les difficultés auxquelles se heurtent les étudiants. Le fait de devoir passer par différentes étapes avant de pouvoir consulter des documents s'avère pénible, sans compter que l'on a tendance à se limiter à une ou deux bases au détriment des autres. Il ne fait aucun doute qu'une interface qui permet d'interroger à partir d'un champ unique l'intégralité des différentes collections se révèle très utile. Le recours au tri et au filtrage après coup semble être une meilleure stratégie que celle qui consiste à proposer dès le départ une interface trop riche.

Afin de s'adapter aux différentes attentes des nouvelles générations, les collections numériques doivent faire l'objet d'une médiation. Les bibliothèques de l'enseignement supérieur se doivent d'offrir des services interactifs et dynamiques d'accès à l'information, avec un système de recherche capable de s'adapter aux besoins des utilisateurs. La tendance en cours auprès des acteurs du secteur est de remplacer les solutions de recherche fédérée, apparues il y a à peine une dizaine d'années, par des outils de recherche unifiée. Cette méthode s'avère gagnante au vu des premières études qui ont pu être menées. Les requêtes n'interrogeant qu'un seul silo, cela permet d'offrir un temps de réponse très cours pour l'affichage des résultats.

Le catalogue doit produire des données utilisables par les hommes, mais aussi par les machines. Les notices doivent pouvoir être exploitées par les programmes afin d'améliorer les services rendus aux usagers. L'intégration de données issues de différentes sources ne se fait pas sans difficultés, nous le savons. Nous espérons que sur ce point, les éditeurs et les fournisseurs parviendront à s'harmoniser. Cela serait bénéfique pour tout le monde. Comme le souligne Hubert Guillaud[97], *« les sites web d'aujourd'hui tirent leurs succès de la réinvention de catalogues de données. Le problème c'est qu'il faut les réinterroger, leur donner un projet, une vision, un but, des objectifs… autres qu'un projet de référencement sans fin »*.

[97] Hubert Guillaud, *Pourquoi avons-nous besoin de catalogues 2.0 ?*, blog « La Feuille », billet publié le 28 avril 2010
http://lafeuille.blog.lemonde.fr/2010/04/28/pourquoi-avons-nous-besoin-de-catalogues-20/

L'arrivée prochaine du web sémantique combinée aux évolutions du web social devrait encore faire évoluer les usages et les outils. Les avancées de la recherche dans les domaines du traitement automatique des langues et de l'intelligence artificielle font que les moteurs de demain seront en mesure d'analyser le sens des mots et de mettre en relation des concepts. Avec la montée en puissance des ordinateurs, on peut imaginer sans mal que les interfaces soient de plus en plus adaptées à nos besoins, capables de réagir efficacement à différents contextes sémantiques, capables aussi de retrouver les informations les plus pertinentes en fonction de nos goûts, personnalisées selon nos contextes individuels. Cela devra se faire au travers de la simplicité, sans quoi on risque de se retrouver une fois de plus avec un faible taux d'utilisation des solutions proposées.

Concernant « la Fusée », avant d'envisager de passer à une solution de recherche globale, il apparaît nécessaire dans un premier temps d'effectuer une analyse des usages des sources actuelles. Cette étude est en cours de réalisation par une stagiaire au sein de la médiathèque, mais nous n'avons pas encore pu profiter des premiers retours. L'analyse des logs de connexion depuis les ordinateurs de SKEMA devrait permettre d'identifier clairement quelles sont les pratiques de recherche des étudiants, tant au niveau des moteurs de recherche proposés par les Geemiks que par les moteurs du web (*Google*, *Google Scholar*, etc.). Ces informations devraient également apporter des indications précieuses sur les types de recherches effectuées (requêtes saisies, nombre de mots employés, durée des recherches, ...). On devrait également pouvoir compter sur les statistiques fournies par l'outil *Google Analytics* installé sur le site de la médiathèque, ainsi que sur les données proposées depuis l'interface d'administration d'*AquaBrowser* (qui ne concerne que le catalogue interne). Ces différents indicateurs permettront d'une part, de repenser éventuellement, en l'améliorant, l'interface actuelle qui donne accès aux différentes bases et, dans un second temps, de guider les choix quant à la suite à donner aux différents abonnements, en partenariat avec le CERAM. Sur le plan des interfaces visuelles, il sera utile de s'interroger sur la manière de guider l'utilisateur jusqu'à la base la plus en adéquation avec ses besoins. Une des solutions pour cela sera de proposer un thésaurus visuel, à la manière de ce que propose la *London Business School*.

L'étape suivante sera la mise en place du nouvel SIBG (*Koha*) qui, tant qu'il ne sera pas prêt à l'emploi, ne permettra pas d'envisager l'intégration d'une solution de recherche unifiée. Pour le choix d'une solution, il sera également utile d'avoir une vision plus claire en ce qui

concerne les autres données externes à la médiathèque qui devront être indexées. Cela passera nécessairement par une communication entre les différents services concernés. D'autre part, comme nous l'avons évoqué dans le dernier chapitre, les solutions de recherche unifiée qui répondent le mieux aux besoins actuels sont encore jeunes et sujettes à améliorations. Les premiers retours d'expérience d'universités françaises devraient aider les futurs clients à choisir le bon produit. Il faudra pour cela attendre 2011, le temps que ces solutions soient déployées et de voir les accords que trouveront les fournisseurs avec les éditeurs francophones.

7 RÉFÉRENCES

AMAR M., MESGUICH V., *Le Web 2.0 en bibliothèques, Quels services ? quels usages ?*, Cercle De La Librairie, 2009

ARSENAULT C., PARÉ F.-X., *Les portails de bibliothèque: nouvelles fonctionnalités, nouveaux défis*, Argus, vol. 34, n°2, automne 2005
https://papyrus.bib.umontreal.ca/jspui/bitstream/1866/136/1/Les%20portails%20de%20bibliothèque%202.pdf

ASSADI H., BEAUDOUIN V., *Comment utilise-t-on les moteurs de recherche sur Internet ?*, Réseaux, n° 116, Lavoisier, pp. 171 à 198, 2002
http://www.cairn.info/load_pdf.php?ID_ARTICLE=RES_116_0171

BALMISSE G., *Classification à facettes, une autre approche de l'organisation de l'information*, KnowledgeNews n°29, janvier 2007
http://www.gillesbalmisse.com/v2/spip.php?article172

BALMISSE, G., *Recherche d'information en entreprise: une question de gouvernance*, in « Usages, usagers et compétences informationnelles au 21è siècle », Lavoisier, 2008

BARRIER G., *Internet, clefs pour la lisibilité - Se former aux nouvelles exigences de l'hypermédia*, ESF éditeur, 2000

BASTIEN J.M.C., SCAPIN D., *Critères ergonomiques pour l'évaluation d'interfaces utilisateurs*, Rapport technique INRIA, n°156, 1993
http://www.webmaestro.gouv.qc.ca/publications/archives/webeducation1998-2004/2000-11/criteres.pdf

BATES M., *The design of browsing and berrypicking techniques for the online search interface*, Online Review, n°13, pp. 407-424, 1989
http://www.scils.rutgers.edu/~tefko/Courses/612/Articles/Bates.pdf

BONNEL N. et al, *Visualisation en recherche d'information*, in « Information et visualisation - Enjeux, recherches et applications », Cépaduès éditions, 2008

BRÉBION P., *Ressources numériques : l'explosion*, Archimag n°235, juin 2010

BRUZA P., DENNIS S., *Query reformulation on the internet: Empirical data and the hyperindex search engine*, 1997
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.44.1331&rep=rep1&type=pdf

BURKE J., *Discovery versus disintermediation: the new reality driven by today's end-user*, VALA 2010, Proquest, 2010
http://www.vala.org.au/vala2010/papers2010/VALA2010_57_Burke_Final.pdf

CARTIER M., *Icône ou pas ?*, in « Les icônes à l'écran », 1990
http://www.michelcartier.com/McArticleB.php3?id_article=382

CALVARY G., *Ingénierie de l'interaction homme-machine: rétrospective et perspectives*, in « Interaction homme-machine et recherche d'information », Hermès Science/Lavoisier, 2002
http://iihm.imag.fr/publs/2002/TSTI.calvary.pdf

CHEVALIER M., *Interface adaptative pour l'aide a la recherche d'information sur le web*, thèse présentée à l'université Paul Sabatier de Toulouse, 2002
http://hal.archives-ouvertes.fr/tel-00350508/

CHEVALLET J.-P., *Les interfaces pour la recherche d'information*, in « Interaction homme-machine et recherche d'information », Hermès Science/Lavoisier, 2002

CHIEZE E., *Reformulation automatique de requêtes par intégration d'éléments syntaxiques dans le cadre du repérage de l'information en français sur le Web*, thèse de doctorat, Université du Québec à Montréal, 2006

CHAUVIN S., *Enjeux et perspectives de la visualisation heuristique pour les catalogues en ligne*, in « Usages et pratiques dans les bibliothèques numériques », Hermes Science Publications / Lavoisier, 2007

COOPER A., *Le mythe de la métaphore*, Visual Basic Programmer's Journal, juin 1995
http://www.multimedialab.be/doc/alan_cooper.htm

DE ROSA C. et al, *College Students' Perceptions of Libraries and Information Resources*, rapport de l'OCLC, décembre 2005
http://www.oclc.org/ca/fr/reports/perceptionscollege.htm

FAVIER L., IHADJADENE M., *Vers des systèmes de découverte et de filtrage d'information documentaire: quelle stratégie faut-il mettre en place ?*, ACSI 2000
http://halshs.archives-ouvertes.fr/docs/00/06/20/84/HTML/

FEKETE J.-D., *Visualisation pour les bibliothèques numériques*, Document numérique, volume 9, février 2006
http://www.cairn.info/revue-document-numerique-2006-2.htm

GAVIN P., *Une nouvelle génération d'OPAC : le Paradis retrouvé ?*, BCU info, n°59, juillet 2008, pp. 3-27
http://www.fr.ch/bcu/n/actuel/BCU_info/info_59.pdf

GLADNEY H. M. et al., *Digital Library: Gross Structure and Requirements, Report from a March 1994 Workshop*, in « Digital Libraries'94 - Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries »
http://atg1.wustl.edu/DL94/paper/fox.html

GRIFFITHS J., BROPHY P., *Student searching behavior and the Web: use of academic resources and Google*, Library Trends, vol. 54, n°4, pp. 539-554, 2005
https://www.ideals.illinois.edu/bitstream/handle/2142/1749/Griffiths539554.pdf

HASCOËT M., *Visualisation d'Information et Interaction*, in « Méthodes avancées pour les systèmes de recherche d'informations », Hermès Science/Lavoisier, 2004

HAWKINS D.T., WAGERS R., *Online bibliographic search strategy development*, Online, vol. 6, n°3, pp. 12-19, mai 1982

HERTZUM M., FROKJAER E., *Browsing and querying in online documentation: A case study of user interfaces and the interaction process*, ACM Transactions on Computer-Human Interaction, vol. 3, n°2, pp. 136-161, 1996
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.86.1028&rep=rep1&type=pdf

HSIEH-YEE I., *Effects of Search Experience and Subject Knowledge on Online Search Behavior : Measuring the Search Tactics of Novice and Experienced Searchers?*, Journal of the American Society for Information Science, n°44, 1993, pp. 161-174
http://www3.interscience.wiley.com/journal/10049852/abstract

HUDON M., *L'utilisation de facettes pour la classification et l'accès aux ressources numériques : proposition et exemple*, in « Usages et pratiques dans les bibliothèques numériques », Hermès Science Publications / Lavoisier, 2007

IHADJADENE M., *La Recherche et la navigation dans un système de recherche d'information grand public*, thèse de doctorat en science de l'information, Université Lyon 1, 1999
http://cat.inist.fr/?aModele=afficheN&cpsidt=203190

IHADJADENE M., *L'analyse des échecs et la surcharge d'informations dans un SRI: examen des tactiques mises en œuvres*, ENSSIB, 1999
http://www.cais-acsi.ca/proceedings/1999/Madjid_1999.pdf

JONES S., MCINNES S., *Graphical query specification and dynamic result previews for a digital library*, in « Proceedings of the 11th annual ACM symposium on User Interface Software and Technology (UIST'98) », november 1998
http://www.cs.waikato.ac.nz/~stevej/Research/PAPERS/uist98.pdf

JOUGUELET S., *Les Learning centres : un modèle international de bibliothèque intégrée à l'enseignement et à la recherche*, décembre 2009, rapport de l'IGB (Inspection générale des bibliothèques)
http://media.enseignementsup-recherche.gouv.fr/file/2009/33/6/Rapport_Learning_Centers_7-12_RV_131336.pdf

KALINICHENKO L. A. et. al., *Digital libraries: advanced methods and technologies, digital collections*, in « All-Russian Scientific Conference, 2002 ». Conference Report Dubna: RCDL, 2002
http://www.dlib.org/dlib/january03/kalinichenko/01kalinichenko.html

KEIM D.A., KRIEGEL H.P., *Possibilities and limits in visualizing large databases*, Visual Database Systems (VDB), 1995, pp 203-214

KWASNIK B. H., *The role of classification in knowledge representation and discovery*. Library Trends, vol. 48, n°1, pp. 22-47, été 1999
http://www.ideals.illinois.edu/bitstream/handle/2142/8263/librarytrendsv48i1d_opt.pdf

LAMOUROUX M., BATTISTI M., *73è congrès de l'IFLA. Bibliothèques du futur : progrès, développement et partenariats*, Documentaliste-Sciences de l'Information, vol. 44, avril 2007
http://www.cairn.info/revue-documentaliste-sciences-de-l-information-2007-4-page-327.htm

LE BESCOND I., *Le portail outil d'intégration des ressources de la bibliothèque hybride : le cas de la Bibliothèque Universitaire de l'Université Libre de Berlin (Allemagne)*, Mémoire d'étude, janvier 2005
http://enssibal.enssib.fr/bibliotheque/documents/dcb/lebescond-vol1.pdf

LEBLOND C., *Un système d'information documentaire à l'université, fédération des ressources et personnalisation des services : de l'idéal du projet à la réalité des usages*, in « Usages et pratiques dans les bibliothèques numériques », Hermes Science Publications / Lavoisier, 2007

LIPPINCOTT J.K., *Net generation students & libraries*, Educause Review, vol. 40, n°2, pp. 56-66, mars 2005
http://net.educause.edu/ir/library/pdf/ERM0523.pdf

MAISONNEUVE M., *Recherche multibases: de nouveaux outils pour accroître l'autonomie des usagers*, Documentaliste, vol. 40, n°3, 2003
http://www.cairn.info/revue-documentaliste-sciences-de-l-information-2003-3-page-214.htm

MAISONNEUVE M., *Les logiciels portails pour bibliothèques et centres de documentation – L'offre d'outils de recherche fédérée et de gestion de contenu*, Tosca Consultants, ADBS éditions, 2007

MAISONNEUVE M., *Le catalogue de la bibliothèque à l'heure du Web 2.0 - Étude des opacs de nouvelle génération*, Tosca Consultants, ADBS éditions, 2008

MAISONNEUVE, M., *Le catalogue à l'heure du Web 2.0 - Présentation de l'offre d'opacs de nouvelle génération*, Tosca Consultants, 4 mars 2009
http://www.toscaconsultants.fr/articles/opacsdenouvellegenerationADBSGrenoble.pdf

MAISONNEUVE M., TOUITOU C., *Une nouvelle famille d'Opac - Navigation à facettes et nuages de mots*, BBF, tome 52, n°6, 2007
http://bbf.enssib.fr/consulter/bbf-2007-06-0012-002.pdf

MARCHIONINI G., *Information seeking in electronic environments*, Cambridge University Press, 1995

MARTIN A., PERRET C, *Les méthodes de recherche documentaire des étudiants de 1ère année de LLCE en début d'année*, Centre d'Innovation Pédagogique et d'Evaluation - Université de Bourgogne, janvier 2009
http://www.enssib.fr/bibliotheque-numerique/notice-48093

MKADMI A., SALEH I., *Bibliothèque numérique et recherche d'informations*, Hermes Science Publications / Lavoisier, 2008

MORIZIO C., *La recherche d'information*, Armand Colin, 2006

MORVILLE P., *Architecture de l'information pour le web (3ème édition)*, O'Reilly, 2007

NAVARRO-PIETRO R., *Cognitive Strategies in Web Searching*, in « Institut of Stanarts and Technology, Proceedings of the 5th Conference on human Factors and the Web », pp. 43-56, juillet 1999
http://zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto

NICHOLAS, D. et al, *Work Package IV: Student Information-Seeking Behaviour in Context - Key findings from CIBER log studies*, University College London, 22 juin 2007
http://www.ucl.ac.uk/infostudies/research/ciber/downloads/GG Work Package IV.pdf

NIELSEN J, *Usability Engineering*, Academic Press, 1993

NIELSEN J., LORANGER H., *Site Web: priorité à la simplicité*, Pearson Education, 2008

O'DAY V. L., JEFFRIES R., *Orienteering in an information landscape: How information seekers get from here to there*, in « Proceedings of ACM/InterCHI'93 », p. 110, 1993
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.90.5480&rep=rep1&type=pdf

PAPY F., CHAUVIN S., *Les usagers et le Visual Catalog - Analyse de 82 questionnaires : octobre 2004 - février 2005*, Rapport de recherches, septembre 2005
http://doc.univ-paris8.fr/fpapy/pubs/QuestionnairesBUsept05.pdf

PAPY F., LEBLOND C., *L'interface de recherche d'information du Visual Catalog : un outil innovant à double détente*, Documentaliste - Science de l'information, vol. 44, n° 4-5, pp. 288-298, 2007
http://www.cairn.info/load_pdf.php?ID_ARTICLE=DOCSI_444_0288

PAPY F., LEBLOND C., *Bibliothèques numériques: la nécessaire médiation*, Communication et langages, n° 161, pp. 37-57, 2009
http://www.necplus.eu/action/displayAbstract?aid=2426076

ROBERT O., *Implantation d'un métamoteur de recherche et d'un résolveur de liens OpenURL en milieu documentaire spécialisé*, Premier congrès des milieux documentaires du Québec, 11 novembre 2009
http://www.asted.org/_uploadedcontent/medias/content_1078_1402.pdf

ROBERTSON S.E., HANCOCK-BEAULIEU M.M., *On the evaluation of IR systems*, Information processing & management, vol. 28, n°4, pp. 457-466, 1992
http://www.soi.city.ac.uk/~ser/papers/ipmeval.pdf

ROWLANDS I. et al, *The Google generation: the information behaviour of the researcher of the future*, ASLIB Proceedings : New information perspectives, vol. 60, n°4, pp. 290-310, 11 janvier 2008
http://www.educnet.education.fr/veille-education-numerique/mai-2010/le-comportement-informationnel-du-chercheur-du-futur

SCHMUTZ B., *Les publics du livre numérique*, Ipsos MediaCT, 29 mars 2010
http://www.ipsos.fr/CanalIpsos/attachment/3103-3-livreNumerique-97.pdf

SEBRECHTS M. et al., *Visualization of search results: a comparative evaluation of text*, 2D and 3D interfaces, ACM SIGIR, Berkley, 1999, pp. 3-10
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.39.4199&rep=rep1&type=pdf

SIMONNOT B., GALLEZOT G., *L'entonnoir : Google sous la loupe des sciences de l'information et de la communication*, C&F Editions, mai 2009

SPINK A. et al, *Searching the web: The public and their queries*, Journal of the American Society for Information Science, vol. 52, n°3, pp. 226-234, 2001
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.23.9800&rep=rep1&type=pdf

STEFANER M., *MACE - Deliverable D7.6 - Report on user interface design and community experiments*, DSpace at Open Universiteit Nederland, 2008
http://dspace.ou.nl/dspace/bitstream/1820/1765/1/Mace%20Deliverable%207.6%20Report%20on%20User%20Interface%20Design%20and%20Community%20Experiments.pdf

TALLENT E., *Metasearching in Boston College Libraries : A Case Study of User Reaction*, New Library World, vol. 105, n°1196/1197, 2004
http://escholarship.bc.edu/cgi/viewcontent.cgi?article=1000&context=library_pubs

TAYLOR A. G., *Introduction to Cataloging and Classification - 9th edition*, Libraries Unlimited, 2000

VAILLANT P., *Sémiotique des langages d'icônes*, Honoré Champion éditeur, 1999

ZAPHIRIS P., MTEI L., *Depth vs Breadth in the Arrangement of Web Links*, 1997
http://otal.umd.edu/SHORE/bs04/