

MULTIPLIER L'ACQUISITION DE BASES DE DONNÉES BIBLIOGRAPHIQUES POUR ENRICHIR LA RECHERCHE DOCUMENTAIRE : MYTHE OU RÉALITÉ ?

L'exemple du domaine linguistique et littéraire

Sara DECOSTER

Responsable scientifique, Université de Liège (ULg) – Bibliothèque ALPHA

Stéphanie SIMON

Responsable scientifique, Université de Liège (ULg) – Bibliothèque ALPHA

Muriel van RUYMBEKE

Directrice, Université de Liège (ULg) – Bibliothèque ALPHA

▪ Cet article propose une analyse approfondie de quatre bases de données bibliographiques sous abonnement disponibles à l'Université de Liège, sur deux plateformes différentes, dans le domaine des études linguistiques et littéraires. Les outils concernés offrent une couverture générale du domaine, au sens où ils ne sont pas spécialisés dans une période chronologique ou une langue particulière. Les quatre bases de données analysées sont *MLA International Bibliography* et *Communication & Mass Media Complete (CMMC)*, accessibles via Ebsco, ainsi que *Francis* et *Linguistics and Language Behavior Abstracts (LLBA)* sur la plateforme ProQuest. L'objectif de l'article est d'évaluer le bénéfice que l'utilisateur peut tirer de ces produits. Afin d'y parvenir, l'analyse suit des axes centrés sur le contenu et la complémentarité des ressources ainsi que sur les fonctionnalités proposées. Les résultats de ce travail peuvent alimenter une réflexion sur l'évolution de la bibliothéconomie et sa place dans le monde de la recherche et de l'enseignement.

▪ Dit artikel geeft een gedetailleerde analyse weer van vier bibliografische databanken m.b.t. letterkunde en linguïstiek uit het abonnementenpakket van de Universiteit de Liège, die toegankelijk zijn via twee verschillende interfaces. Het gaat om algemene zoekinstrumenten, in die zin dat zij noch gespecialiseerd zijn in een specifiek tijdvak, noch in een bepaalde taal. De bestudeerde databanken zijn enerzijds *MLA International Bibliography* en *Communication & Mass Media Complete (CMMC)*, beschikbaar via Ebsco, en anderzijds *Francis* en *Linguistics and Language Behavior Abstracts (LLBA)* via het ProQuest-platform. Het doel van dit artikel bestaat erin te onderzoeken welk voordeel de gebruiker kan halen uit deze producten. Met dit oogmerk wordt zowel een inhoudsgerichte onderzoekslijn gevolgd, die ook de complementariteit van de verschillende hulpmiddelen meet, als een functioneel gerichte onderzoekslijn. De resultaten van deze studie kunnen nuttig zijn om na te denken over de evolutie van de bibliotheekwetenschap en haar plaats in de onderzoeks- en onderwijswereld.

Dans le vaste marché des produits documentaires, la décision d'acquérir telle ou telle ressource plutôt que telle ou telle autre dépend de plusieurs facteurs¹. Le prix ou l'acquisition en consortium entrent notamment en ligne de compte mais ils ne constituent pas les critères décisifs. En effet, les produits seront plutôt sélectionnés en fonction de leur contenu et des fonctionnalités qu'ils proposent. Dans cet ordre d'idées, cet article propose une analyse approfondie de quatre bases de données bibliographiques sous abonnement à l'Université de Liège, sur deux plateformes différentes, dans le domaine des études linguistiques et littéraires. Grandes références dans ces disciplines, les outils concernés offrent une couverture générale du domaine, au sens où ils ne sont pas spécialisés dans une période chronologique ou une langue particulière. L'objectif sera d'évaluer le bénéfice que l'utilisateur peut tirer de ces produits. L'étude se déclinera selon des axes centrés sur le contenu et la complémentarité des ressources et sur les fonctionnalités proposées. En effet, un produit intéressant est également un produit

maniable, d'où l'importance de l'ergonomie. Les résultats de ce travail peuvent alimenter une réflexion sur l'évolution de la bibliothéconomie et sa place dans le monde de la recherche et de l'enseignement.

La première bibliographie analysée est *MLA International Bibliography*² sur la plateforme d'Ebsco, une référence dans le domaine littéraire, également utile dans les disciplines connexes comme la linguistique et les arts de la scène³. La deuxième base de données d'Ebsco est *Communication & Mass Media Complete (CMMC)*, qui couvre le domaine de la communication au sens large, donc également sous ses aspects langagiers. Sur la deuxième plateforme, ProQuest, l'examen porte sur l'outil en sciences du langage *Linguistics and Language Behavior Abstracts (LLBA)* et sur *Francis*. Cette dernière base multidisciplinaire couvre la littérature et la linguistique, mais aussi d'autres domaines de sciences humaines, tels que l'histoire, la sociologie et la psychologie. Il s'agit de quatre produits qui se partagent désormais le marché avec

d'autres bases comme la nouvelle version en ligne de la *Romanische Bibliographie* (De Gruyter) et la *Linguistic Bibliography Online* (Brill), récemment réélaboree, qui se distingue par la grande attention portée aux langues minoritaires ou extérieures au territoire occidental. Ces bibliographies sont présentes à l'Université de Liège sous format papier.

Comparaison des contenus proposés par les quatre bases de données

Pour répondre à la question de l'originalité, de la complémentarité ou de la concurrence entre les quatre bases de données étudiées, l'analyse des contenus proposés par chacune est essentielle.

Nombre de références et typologie de documents

Le premier critère de comparaison sera quantitatif. Les quatre bases de données sont loin d'être égales en nombre de documents référencés selon le premier tableau.

Tab. 1 : Nombre de références et de titres dépouillés dans les quatre bases de données.

Situation au 25/07/2012	MLA	Francis	CMMC	LLBA
Nombre de références	2.452.290	1.972.548	580.486	512.116
Nombre de titres de périodiques dépouillés	6.735	2.706	808	3.598

Cette différence dans le nombre de références globales s'explique en partie par la couverture thématique des bases : *Francis* et *MLA* explorent des domaines plus vastes que les deux autres qui se veulent davantage ciblées sur certains aspects de la linguistique et des sciences humaines. Si le nombre de références fait la richesse d'une base de données, il est clair que la *MLA* et *Francis* présentent un certain atout. Toutefois, un chercheur spécialisé dans une discipline particulière aura probablement besoin de davantage de références précises ou techniques. Dans ce cas, *Francis* sera probablement trop généraliste en comparaison avec les trois autres bases, plus ciblées au plan thématique.

La typologie des documents indexés par les bases de données peut également être un point de comparaison intéressant. Le second tableau présente le pourcentage de chaque type de documents référencé

Tab. 2 : Typologie des documents rencontrés dans les quatre bases de données

Typologie de documents	MLA	Francis	CMMC	LLBA
Articles	63,4 %	87,9 %	80,8 %	73,1 %
Livres	30,9 %	3,3 %	1,3 %	4,5 %
Thèses	5,7 %	0,9 %	-	7,6 %
Actes de colloque et conférences	-	5,9 %	1,2 %	0,1 %
Comptes rendus de livres	-	-	12,2 %	14,6 %
Documents divers	< 0,1 %	2 %	4,5 %	< 0,1 %

dans chacune des quatre bases.

Les articles de revue constituent les principales références des quatre bases. *CMMC* et *LLBA* se spécialisent ensuite dans l'indexation de comptes rendus de livres. La *MLA* référence également un grand nombre de livres et chapitres de livres. Enfin, *Francis* indexe un grand nombre d'actes de conférences. Ces principaux types de documents sont complétés de façon plus anecdotique par des documents tels que des thèses, documents légaux, éditoriaux, livres, bibliographies, comptes rendus, etc., en diverses proportions.

Couverture exhaustive ou sélective des périodiques

Principales références, les articles sont extraits de périodiques dont le nombre et la couverture varient selon les bases. Cette couverture peut être exhaustive : chaque article d'un périodique fait l'objet d'une référence dans la base de données. C'est le cas dans *CMMC* et la *MLA*. Pour l'utilisateur, une recherche dans ces deux bases lui donne ainsi la certitude de couvrir l'intégralité du contenu des revues dépouillées par ces mêmes bases.

Dans d'autres cas, la couverture n'est que sélective. Un éditeur scientifique dépouille les périodiques et n'insère dans la base que les références jugées pertinentes en fonction de certains critères, parmi lesquels on peut supposer la prise en compte de la proximité thématique et géographique avec les domaines couverts annoncés par l'éditeur. *Francis*, dont le contenu est géré par l'Institut de l'Information Scientifique et Technique (INIST), est un cas illustrant ce principe de couverture sélective. Cet institut base sa sélection sur des critères thématiques et scientifiques qui ne sont cependant pas accessibles au grand public.

Enfin, la *LLBA* offre un mélange des deux politiques : une minorité de titres sont dépouillés et

référéncés intégralement⁴, tandis que la majorité ne fait l'objet que d'une indexation sélective. Le principe est intéressant mais pas transparent pour l'utilisateur qui ne peut savoir quels sont les titres concernés et avoir donc l'impression, erronée, d'avoir recherché dans l'intégralité de ces revues.

Couverture temporelle : état de l'indexation et de la publication

La couverture temporelle des titres est fort inégale d'un périodique à l'autre, et d'une base de données à l'autre. Ces données sont par ailleurs difficilement accessibles⁵. La seule base de données pour laquelle ces données sont accessibles en ligne est *CMMC*⁶. Parmi l'ensemble des 808 titres, 118 sont indexés depuis moins de trois ans⁷. 415 titres font l'objet d'un dépouillement pour les années antérieures à l'an 2000 et 183 titres seulement avant 1990. Cette couverture temporelle variable est due en partie aux dates de parution des revues en elles-mêmes, mais c'est loin d'être systématique.

Enfin, l'"actualité" de l'indexation diffère grandement d'une base à l'autre. En effet, tous les titres de périodiques ne sont pas activement indexés : les titres sont toujours présents dans la liste des périodiques recensés dans la base, mais les nouveaux fascicules ne sont plus dépouillés ni référencés⁸. Cet aspect peut se révéler important, puisque de l'activité de l'indexation dépend également la capacité d'accroissement d'une base de données. Or, les chiffres obtenus posent question⁹. Dans *Francis*, sur un échantillon de 150 revues, 96 % sont toujours en cours de parution. Néanmoins, cela ne signifie pas qu'elles soient toujours activement indexées. Les autres bases de données ont pu être analysées plus précisément. Dans la *CMMC* et la *MLA*, les titres activement indexés sont toujours majoritaires : respectivement 65 % et 72 % de l'ensemble des titres. Dans la *LLBA* par contre, plus de 62 % de la liste des titres fournies par ProQuest ne sont plus activement indexés. Seuls 38 % de leurs titres permettent donc encore d'enrichir régulièrement la base de données.

L'état de l'indexation peut être lié aux dates de parution des revues en elles-mêmes. La base *LLBA* pourrait ainsi s'être concentrée sur des revues plus anciennes qui ne sont plus guère publiées aujourd'hui, ce qui expliquerait cette proportion de titres à indexation clôturée. Nous avons analysé un échantillon de titres dont le dépouillement a cessé pour chaque base de données. Le but a été de vérifier, pour cette liste uniquement, l'état de publication du périodique à l'heure actuelle. La liste des titres dépouillés dans *Francis* ne nous permet pas de tirer de

conclusions, pour les mêmes raisons qu'évoquées précédemment. Par contre, pour *CMMC* et *MLA*, l'état d'activité de l'indexation semble correspondre en grande partie à l'état de publication des titres. Dans *CMMC*, sur un échantillon précis de titres à l'indexation arrêtée¹⁰, un peu plus de 85 % des titres ne sont plus publiés, contre 15 % toujours en cours de parution. Dans *MLA*, sur un échantillon plus restreint¹¹, l'ensemble des titres n'était plus publié. Enfin, l'hypothèse de la présence de davantage de titres anciens dans la *LLBA* ne tient pas la route. Sur un échantillon de 150 revues à indexation arrêtée, 80 titres, soit plus de la moitié, sont toujours en cours de parution à l'heure actuelle. Autrement dit, la fin de parution des revues ne peut expliquer qu'une minorité de cas dans cette base de données. Notons par ailleurs que les chiffres précédents ne sont qu'un aperçu du problème puisqu'il faudrait encore y ajouter les revues arrêtées dont la date de fin de parution ne correspond pas non plus à la date de fin d'indexation. Les données proposées à ce propos par les deux sociétés sont de nouveau difficilement accessibles à l'utilisateur¹².

Concrètement, il est facile de penser que l'utilisateur qui interroge ces bases sera toujours plus content d'avoir une réponse, même incomplète, que de devoir dépouiller les mêmes revues manuellement. C'est exact mais simplificateur. Le problème principal est le manque de transparence vis-à-vis de l'utilisateur : la liste de milliers de titres proposée par chaque fournisseur d'accès laisse entendre à l'utilisateur qu'il dépouille effectivement l'ensemble de ces titres. C'est pourtant loin d'être le cas, puisque sa recherche parmi les quatre bases de données l'amène à laisser de côté un important nombre de volumes et d'articles qui ne sont pas indexés. Il est déjà compliqué pour le professionnel de l'information d'arriver à la liste des titres et leur couverture (quand elle est proposée). Comment penser que l'utilisateur pourra y parvenir sans aide particulière ? Cette indexation partielle se montrera particulièrement dangereuse pour les usagers débutants, insuffisamment méfiants ou mal informés, qui se contenteront des résultats fournis par les bases de données sans poursuivre manuellement leur enquête. Ils risqueront alors de passer à côté de références importantes et donc d'affaiblir leur travail.

Contenu linguistique

Les recherches en sciences humaines en général sont souvent influencées par le lieu où elles voient le jour. En Belgique, nos utilisateurs, francophones, néerlandophones ou germanophones, se créent des sujets d'étude souvent liés à leur environnement direct : un historien liégeois étudiera le Moyen Âge liégeois ou belge, un historien de l'art aura davantage de chances d'analyser une œuvre belge également. Il importe donc aux professionnels de l'information de mettre à disposition des utilisateurs des ressources permettant un accès à cette documentation locale ou régionale. Dans le domaine littéraire et linguistique plus précisément, si cet ancrage local est moins marqué, la documentation de ce premier type ne peut être écartée. Deuxièmement, une grande partie de la littérature scientifique internationale est véhiculée dans des langues largement usitées : l'anglais, le français ou l'allemand. Il importe donc également de mettre à disposition de l'utilisateur la littérature scientifique internationale disponible sur ses sujets de prédilection, même locaux. Enfin, les études en sciences humaines s'intéressent aussi à des sujets internationaux, plus éloignés. Pour ces études précises, l'indexation de quelques documents de référence peut être précieuse aux utilisateurs.

Que constater à ce sujet à partir des documents extraits des quatre bases étudiées ? À partir des formulaires de recherche avancée des plateformes Ebsco et ProQuest, il a été possible d'examiner la proportion de documents référencés dans chaque langue (voir tab. 3).

Concernant la mise à disposition d'une littérature scientifique dans la langue maternelle des utilisateurs de nos bibliothèques¹³, *Francis* offre le panel linguistique le plus large. À l'inverse, *CMMC* donne la priorité à la littérature anglophone. À propos de la littérature scientifique internationale, le tableau est également éclairant. Outre l'anglais, le français et l'allemand sont les langues les plus fréquemment rencontrées. Vient ensuite l'espagnol, l'italien et le russe. Le néerlandais est globalement assez peu représenté. D'autres langues, très diverses, se retrouvent enfin en très faibles proportions dans les quatre bases de données.

Sur le plan linguistique, les quatre ressources proposent une couverture suffisamment étendue,

Tab. 3 : Contenu linguistique des quatre bases de données.

Langue des documents référencés	MLA	Francis	CMMC	LLBA
Anglais	59,6 %	57,8 %	98,2 %	71,6 %
Français	8,7 %	26,2 %	0,3 %	7,0 %
Allemand	8,0 %	5,9 %	0,5 %	6,5 %
Espagnol	5,8 %	3,1 %	0,8 %	2,6 %
Italien	3,4 %	2,4 %	0,1 %	1,5 %
Russe	2,8 %	1,5 %	0	2,1 %
Néerlandais	0,8 %	0,4 %	< 0,01 %	0,8 %
Nombre d'autres langues représentées	71	70	8	53

plus riche pour les trois premières que pour *CMMC*. Dans notre université, chacune des quatre bases de données semble donc rencontrer nos exigences. Il faut cependant souligner que cette satisfaction est également liée à notre localisation géographique au carrefour des mondes francophone, néerlandophone et germanophone et à l'utilisation forte de l'anglais dans les milieux universitaires.

Degré de recouvrement des périodiques couverts

Les ressources analysées dépouillent ensemble plusieurs milliers de titres de périodiques. Parmi ceux-ci, il est inévitable de trouver des recouvrements, autrement dit, des titres dépouillés par plusieurs bases simultanément. En évaluer le nombre est nécessaire lorsqu'il s'agit de juger de l'originalité du contenu des bases de données proposées aux utilisateurs. Pour des raisons de faisabilité, afin d'évaluer l'unicité des titres dépouillés, nous avons basé notre analyse sur un échantillon de périodiques¹⁴ pour lequel nous aboutissons à quelques conclusions illustrées graphiquement (voir fig. 1).

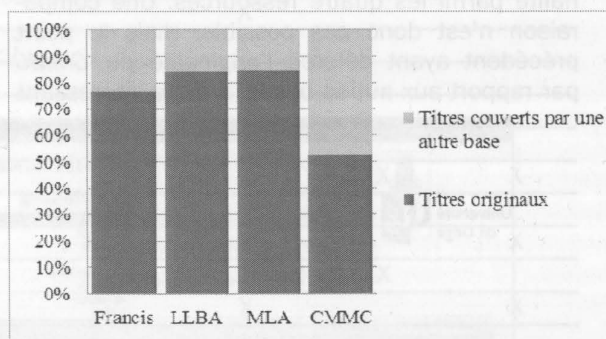


Fig. 1 : Proportion de titres originaux par rapport à l'ensemble des titres dépouillés dans chaque base.

La proportion de titres uniques est très variable selon la ressource. Alors que *LLBA* et *MLA* référencent plus de 80 % de titres uniques, la proportion diminue jusqu'à 52 % seulement dans *CMMC*. Le contenu de la base réellement original serait ainsi réduit à la moitié de ses références, si

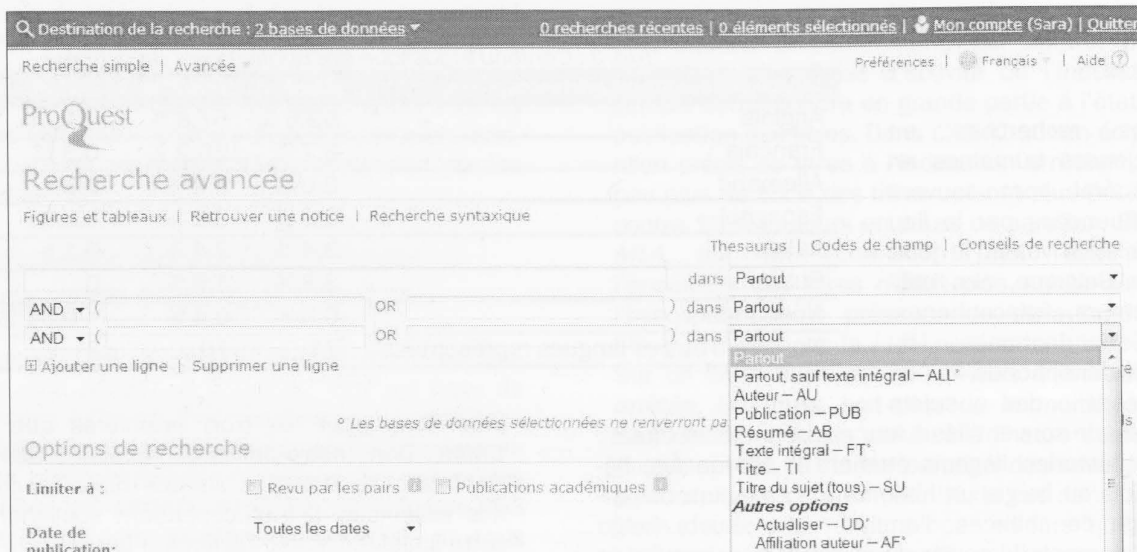


Fig.3 : L'interface ProQuest.

l'on ne tenait compte que du critère des recouvrements pour en évaluer la richesse.

Au niveau des titres présents plusieurs fois, il est intéressant d'observer les relations entre bases de données. Ainsi, pour *Francis*, la *LLBA* et *CMMC*, on constate que la majorité des titres recouverts sont en fait dépouillés également par la *MLA*. Cela pourrait s'expliquer assez simplement par le nombre de titres plus importants dépouillés par cette base de données.

Enrichissement des bases de données par la fourniture d'articles en texte intégral

Il reste un critère d'évaluation de l'originalité de contenu des quatre bases de données à envisager : l'accessibilité au texte intégral des articles référencés. Seule *CMMC* propose cette fonctionnalité parmi les quatre ressources. Une comparaison n'est donc pas possible, mais le point précédent ayant déformé l'originalité de *CMMC* par rapport aux autres bases, il reste intéressant

de pousser plus loin l'analyse.

Sur l'ensemble des documents référencés dans la base de données¹⁵, le nombre de documents en texte intégral disponibles était de 363.722. Ce nombre est donc équivalent à 62,7 % de son contenu. L'intérêt de *CMMC* est peut-être là. Certes, elle compte près d'une moitié de titres constituant des doublons par rapport aux autres bases de données. Néanmoins, elle est également la seule base de données fournissant un accès direct à plus de la moitié de ses références. Alors qu'une recherche dans la *MLA*, *LLBA* ou *Francis* nécessite toujours un passage par un résolveur de liens, un site extérieur ou par une bibliothèque, une recherche dans *CMMC* fournit un contenu direct à l'utilisateur.

Recherche et indexation

Les deux plateformes hébergeant les quatre bases comparées ont entrepris un effort d'homogénéisation formelle. Elles possèdent

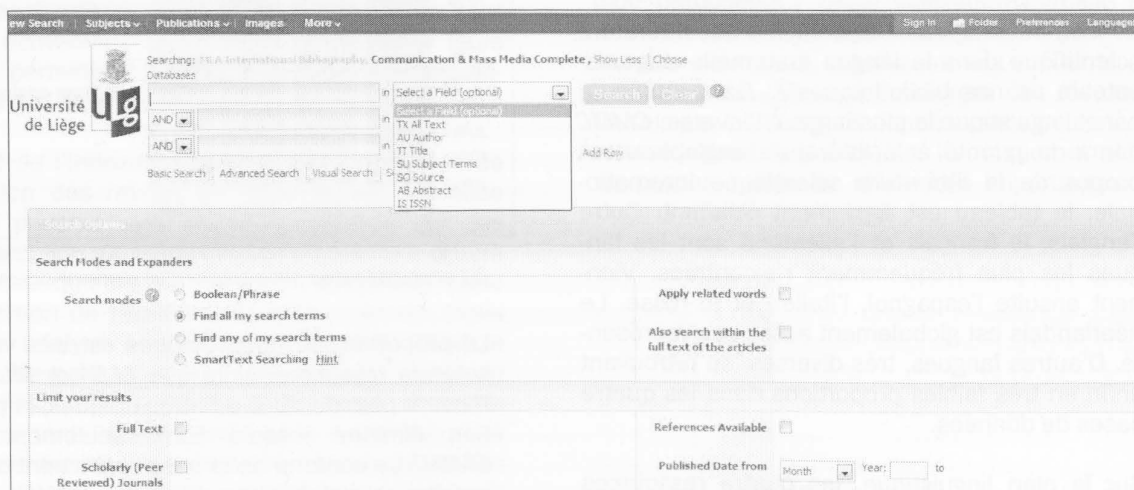


Fig.2 : L'interface EBSCOhost.

chacune leur identité visuelle propre et adoptent un mode de fonctionnement commun pour les différentes bases qu'elles hébergent ; il est d'ailleurs possible d'interroger en une seule recherche différentes bases d'une même plateforme. Malgré ce lissage des interfaces utilisateurs, l'exploitation des quatre outils a révélé un certain nombre de disparités.

Recherche simple ou avancée ?

Dans la plupart des outils de recherche – c'est même le cas de *Google* – plusieurs types de recherche sont disponibles : une recherche très basique ou une recherche avancée permettant de différencier plusieurs critères. Les deux plateformes examinées ici ne dérogent pas à la règle. Soulignons cependant d'emblée que la recherche de base qu'elles proposent ne souffre pas de la même opacité que l'interface de *Google* de base, qui impose des préférences linguistiques et l'opérateur AND, sans que ces choix ne soient clairement exprimés. ProQuest explique que l'interrogation se fait avec l'opérateur implicite AND, tandis que, dans le cas d'Ebsco, c'est à l'utilisateur de définir les rapports entre les différents critères qu'il combine. Sur les deux plateformes, c'est cependant la recherche avancée qui constitue la recherche par défaut à

développée, au sens où elle permet de combiner horizontalement plusieurs termes de recherche avec l'opérateur booléen OR. Cette construction est parfaitement adaptée aux cas de synonymie et garantit une plus grande intelligibilité de l'ordre des opérations.

Avec quels critères ?

Les champs interrogeables sont évidemment limités aux possibilités offertes par chaque base de données interrogée. Au niveau des notices individuelles, la description possède, d'une manière générale, un degré de complétude suffisamment élevé pour que les recherches basiques ne soient pas entravées. Des problèmes peuvent toutefois surgir. Par exemple, dans *MLA*, nous avons relevé des notices anciennes (antérieures à 1985) sans indexation par sujet. D'une manière générale, aucune indexation n'est infaillible¹⁶. D'ailleurs, l'indexation des notices anciennes ne correspond plus nécessairement aux catégorisations actuelles¹⁷. En effet, dans les différentes bases, tous les détails ne sont pas systématiquement disponibles pour les recherches plus élaborées (cf. infra). Un aperçu des principaux critères se trouve dans le tableau 4, qui n'a cependant pas de vocation d'exhaustivité.

Tab. 4 : Critères de recherche disponibles dans les quatre ressources.

Critère de recherche	MLA	Francis	CMMC	LLBA
Tous les champs	X	X	X	X
Auteur	X	X	X	X
Titre	X	X	X	X
Publication (titre de périodique, titre du recueil dans <i>MLA</i>)	X	X	X	X
Sujets (classificateurs, fournis par l'auteur,...)	X	X	X	X
Date	X			
ISSN/ISBN	X		X	
Titre original / autre titre		X		
Autres données bibliographiques (éditeur, collection...)	X	X		
Infos publication, conférence	X	X		
Langue de la publication	X			
Affiliation de l'auteur		X		
Résumé	X	X	X	X
Table des matières	X			
Texte intégral				X
Champs spécialisés ¹	X		X	
Tags utilisateur (cf. infra)		X		X

l'Université de Liège. D'un degré de précision plus élevé, cette recherche permet d'énoncer explicitement, de manière claire et détaillée, les modalités d'interrogation des données. Elle fonctionne toujours sur base de la sélection d'un opérateur booléen, qui détermine comment se combinent les différents critères, cherchés dans les champs choisis par l'internaute. Nous avons remarqué que l'interface de ProQuest est plus

LLBA n'offre que les critères les plus basiques, qui sont néanmoins les plus fréquents. Les autres bases de données proposent une myriade de champs différents à interroger, mais l'utilisateur s'interroge parfois sur leur utilité. La recherche dans le texte intégral, présent à l'écran de recherche pour *LLBA*, ne permet pas d'effectuer une recherche concluante, étant donné que *LLBA* n'inclut pas de recherche "full text". La recherche

en texte intégral n'est pas non plus disponible pour *CMMC*, pourtant le seul produit à réellement fournir un accès aux articles mêmes.

Une remarque semblable peut être formulée concernant la recherche dans les tables de matières et dans les résumés proposée par *MLA*, étant donné la quantité relativement restreinte de notices pour lesquelles cette information se trouve incluse dans la base. *Francis* propose une recherche sur base de l'affiliation des auteurs, mais cette information manque dans beaucoup de descriptions individuelles. De même, *CMMC* autorise la recherche par zone géographique, mais cette information n'est pas systématiquement mentionnée dans les fiches où ce critère peut avoir une pertinence. *MLA* et *CMMC* présentent des zones de recherche adaptées à leurs disciplines, dont certaines offrent une réelle valeur ajoutée. Tel est notamment le cas du critère "primary subject work" dans *MLA* ou, dans *CMMC*, celui de "company entity", c'est-à-dire les entreprises. D'un autre côté, il est parfois difficile de comprendre sous quelle forme les informations doivent être introduites dans la case de saisie. Ainsi, l'utilisateur doit bien se rendre compte que les prénoms des personnes ne sont généralement pas mentionnés dans *Francis*, ni dans la zone "auteurs", ni dans la zone "sujets".

Ce dernier champ pose également des problèmes d'équivalence dans les traductions (cf. infra). Dans les bases accessibles par ProQuest, la différence entre les "sujets" et les "identificateurs" n'est pas claire¹⁸. Il est légitime de supposer que cette différence s'explique par la distinction que l'on observait sur l'ancienne plateforme CSA entre le vocabulaire contrôlé et les simples mots-clés, cette spécificité reflétant simplement une évolution récente des concepts. L'information n'est toutefois pas fournie dans les pages d'aide. Dans *MLA*, la signification de "subject literature", par exemple, est assez obscure à première vue. Il s'agit, en réalité, d'une subdivision géographique et non linguistique. En d'autres termes, ce critère permet de séparer la littérature argentine de la littérature bolivienne, mais n'établit pas le lien avec la langue espagnole. Or, cette indexation s'avère mal adaptée aux cas plus complexes qu'illustrent les auteurs belges : Émile Verhaeren appartient tantôt à la littérature belge, tantôt à la littérature française. Le classement est aussi disparate pour Hugo Claus, qui se retrouve sous "Netherlandic literature", "Belgian literature" et "Flemish literature". En effet, toute indexation comporte toujours une part de subjectivité et est influencée par les sensibilités et les compétences de la personne qui l'effectue. Des abstracts sont présents dans *LLBA*, *Francis* et *CMMC*¹⁹. En tant que condensés

de contenu, les résumés autorisent des recherches très pertinentes.

C'est pourquoi nous avons calculé le pourcentage de notices présentant un résumé sur base d'un échantillon de 400 notices choisies aléatoirement²⁰.

Tab. 5 : Pourcentage de notices faisant l'objet d'un résumé et subdivision linguistique de ceux-ci.

	<i>MLA</i>	<i>Francis</i>	<i>CMMC</i>	<i>LLBA</i>
Résumés présents	3 %	86 %	99 %	95 %

La plupart des résumés sont écrits en langue anglaise. *Francis* propose cependant une proportion importante de résumés en français. À travers les différentes bases, les résumés sont d'une longueur très variable et s'avèrent parfois, dans les faits, très réduits. Ajoutons enfin que certaines informations sont présentes dans les différentes bases, mais ne correspondent pas nécessairement à un critère de recherche spécifique. *CMMC* mentionne l'affiliation des auteurs, sans qu'un critère spécial ne soit prévu. Ce choix peut se justifier par le fait que l'appartenance des auteurs n'est pas toujours connue. L'affiliation est interrogée dans une recherche portant sur l'ensemble des champs, qui génère cependant un bruit important.

Les index

Les plateformes ProQuest et Ebsco utilisent les index de manières assez différentes. Alors que les index s'intègrent dans l'interface de recherche chez ProQuest, ils sont exploités de manière indépendante chez Ebsco. Chez ProQuest, l'introduction d'un critère de recherche dans une zone spécifique mène à l'apparition d'un lien permettant de consulter l'index concerné, soit en parcourant une liste alphabétique, soit en interrogeant les mots inclus. Les résultats de cette opération alimentent directement l'écran de recherche. La même remarque vaut pour le thésaurus, qui s'insère tout aussi harmonieusement dans la construction de la recherche avancée. Les index et le thésaurus bénéficient donc d'une meilleure visibilité chez ProQuest. Par contre, force est de constater que certaines notions ne couvrent pas exactement les mêmes réalités selon qu'elles sont utilisées dans les index ou comme critère dans l'écran de recherche de l'interface. En effet, les éditeurs scientifiques sont repris comme "auteurs" lors d'une recherche de base, tandis qu'ils ne sont évidemment pas indexés comme auteur dans l'index. La sélection dans l'index mènera donc à une recherche plus précise. Si les deux modalités de recherche ne sont pas équivalentes, l'utilisateur devrait être informé sur ce sujet.

Nous avons estimé que la combinaison des index et de propositions de saisie semi-automatiques, que l'utilisateur peut désactiver, crée une certaine confusion. En effet, la saisie semi-automatique ne se base pas sur les index et ne tient pas non plus compte du poids des termes concernés au sein de la base de données utilisée. Elle se construit au fur et à mesure en fonction de la fréquence des critères utilisés par l'ensemble des internautes actifs sur la plateforme, toutes bases confondues. D'autre part, la saisie semi-automatique dépend également de contraintes techniques comme le choix du navigateur. Au-delà des contingences de l'informatique, il est difficile d'admettre que le nombre de saisies puisse guider de manière adéquate la construction d'une équation, et ce d'autant plus qu'aucune particularisation disciplinaire n'est prévue. Étant donné la spécialisation de la recherche universitaire, des résultats de recherche issus de domaines totalement étrangers à celui du chercheur ne sont pas très pertinents. Cette remarque vaut tout particulièrement pour certaines zones d'information, les auteurs, par exemple.

Nous avons particulièrement apprécié, dans *MLA*, la fonctionnalité "names as subjects". Bien plus qu'une simple liste alphabétique de noms d'artistes ou de penseurs, cet outil est conçu sur le principe d'une base d'autorités. Des renvois entre les diverses désignations d'une même personne physique garantissent une recherche complète et univoque. Ce système permet également d'isoler les homonymes. Sur base d'informations comme la nationalité et les dates associées à un personnage, l'internaute peut choisir l'entrée appropriée et l'utiliser comme base de la recherche. Ainsi, il devient possible de sélectionner les notices concernant le poète cubain José María de Heredia sans être dérangé par les références à son homonyme français plus célèbre. Les autres index présents dans les deux bases de données hébergées par la plateforme Ebsco se présentent comme une fonctionnalité avancée, que l'internaute peut consulter pour obtenir des informations plus approfondies. Les index permettent notamment de se renseigner sur les champs proposés dans la recherche multicritères. Les index fonctionnent en effet comme des listes de mots-clés, comportant une indication du nombre d'occurrences dans la base. De fait, les index se singularisent pour chaque base de données individuelle, l'index des auteurs de *MLA* n'étant pas égal à celui de *CMMC*. Le même constat vaut pour *Francis* et *LLBA*.

Dans les différents index des bases de données Ebsco, l'internaute peut également sélectionner les termes les plus intéressants pour sa recher-

che, en les combinant par l'intermédiaire d'un opérateur booléen. Il est possible d'exploiter plusieurs index à la fois et de construire des équations extrêmement compliquées. Le même degré de complexité peut être atteint sur les interfaces ProQuest, grâce à la recherche syntaxique. Celle-ci est proposée sur une interface très conviviale, qui permet néanmoins une grande formalisation et une précision exceptionnelle, grâce à une myriade de champs différents. Les champs ne sont toutefois pas toujours tous pertinents. Ainsi, le système propose les sujets *MeSH*, le thésaurus internationalement utilisé dans le domaine biomédical. Or, ni *Francis*, ni *LLBA* ne font appel à ce type d'indexation. Logiquement, l'internaute n'obtient donc aucun résultat en utilisant cet index pour une recherche sur la notion "speech delay", issue de *MeSH*. Pourtant, *Francis* comprend des références inventoriées sous le sujet "speech delay", dont l'appartenance au *MeSH* n'est cependant pas reconnue par le système. Somme toute, la présence d'index non activés pour la base interrogée est susceptible d'induire l'utilisateur en erreur.

Fonctionnalités avancées et ergonomie

Outre les modalités de recherche susmentionnées, les différentes bases de données donnent accès à d'autres fonctionnalités, offrant soit un surplus d'information, soit une autre forme de présentation.

Les fonctionnalités avancées

En règle générale, une recherche doit être réajustée conformément aux attentes exactes du chercheur. C'est pourquoi l'utilisateur peut, en amont, imposer des limites à la source, pour interroger des sous-ensembles spécifiques des différentes bases de données, ou au contraire interroger plusieurs bases en même temps. Ensuite, les résultats conduisent, en aval, à une révision des critères, voire même à un léger réajustement du sujet de recherche. D'où l'importance du système de faceting, disponible dans de nombreuses bases de données. Les fouilles dans les références bibliographiques permettent d'identifier les grands spécialistes d'une thématique et de discerner les différentes orientations scientifiques. À ce sujet, *CMMC* inclut aussi un répertoire de chercheurs. Dans la même optique, l'hypernavigation est en général prévue pour des champs comme les titres de périodiques. Dans les produits Ebsco analysés, les informations concernant les périodiques dépouillés s'affichent à l'intérieur de la base elle-même. Elles peuvent

donc être consultées parallèlement à la recherche bibliographique et s'intégrer à la recherche.

Par ailleurs, les différentes plateformes permettent de sélectionner une tranche de résultats pour une exploitation ultérieure. Il est toujours possible d'appliquer un filtre de titre de périodique, de typologie documentaire ou par sujet. D'autres filtres sont spécifiques à certaines bases de données. Ainsi, la géographie et le raffinement par entreprise ne sont utilisés que pour *CMMC*. D'une manière générale, le paramétrage des filtres est très riche dans l'interface ProQuest, mais malheureusement toutes les possibilités ne s'appliquent pas à l'ensemble des bases. Cette interface permet néanmoins de choisir un résultat jugé particulièrement pertinent et de cocher plusieurs descripteurs associés, qui serviront de base à une nouvelle recherche. D'un autre côté, les facettes pour les matières sont particulièrement riches dans les interfaces d'Ebsco. Un historique de recherche est présent sur les différentes interfaces. L'utilisateur peut visualiser les résultats obtenus antérieurement et créer des alertes ou combiner plusieurs recherches moyennant un opérateur booléen. Celui-ci doit être saisi manuellement sur la plateforme de ProQuest, tandis que chez Ebsco, il faut cocher les recherches concernées dans une liste.

Notons également que, tant *LLBA et Francis* que *CMMC* utilisent avantageusement les bibliographies des articles cités. Dans *CMMC*, la recherche de documents citant une référence particulière relève d'une fonctionnalité autonome. ProQuest joue là de nouveau la carte de l'intégration : le cas échéant, le système propose des liens vers les bibliographies rattachées aux notices ou, inversement, vers les documents citant ces dernières, voire même vers les documents partageant des références. Bien que de telles fonctionnalités constituent une facilité précieuse pour le chercheur, il importe de mesurer la portée exacte de ce type d'outil : aucune base de données ne suit le panorama complet de l'imbrication des références. En effet, les citations proviennent forcément de notices appartenant à un ensemble référentiel sujet à recherche. Si le système repère des références extérieures à la base, les bibliographies des documents concernés ne sont jamais interrogées. Des réserves s'imposent également pour le repérage de documents similaires, qui est prévu dans ProQuest. En effet, certains rapprochements, basés sur des critères informatiques parfaitement objectifs, s'avèrent peu pertinents.

Une fonctionnalité intéressante à mentionner concerne le traitement des images, tableaux et graphiques. Grâce au "deep indexing", ces données sont réellement sujettes à recherche dans

LLBA et *CMMC*. Chez ProQuest, les images peuvent être insérées dans une présentation PowerPoint par un simple clic, le système ajoutant automatiquement la référence bibliographique appropriée. Ebsco fait attention au droit d'auteur en proposant un lien explicatif dans chaque notice.

Finalement, les deux interfaces proposent des modes de recherche plutôt additionnels. Dans *MLA*, une recherche visuelle organise les résultats d'une manière très graphique, tandis que ProQuest a développé un formulaire de recherche structuré autour des mêmes champs que les formats de citation permettant une vérification rapide de notices individuelles.

L'ergonomie

Les relations entre homme et machine font l'objet de l'ergonomie : tout outil existe par l'usage que l'être humain en fait. L'utilisation doit donc être sûre, confortable et efficace²¹. Dans une base de données, l'ergonomie se manifeste dans la commodité de la navigation, la lisibilité des données et l'information à l'intention de l'utilisateur, pour que celui-ci puisse faire un usage correct du système. Tous ces éléments contribuent à l'expérience globale de l'utilisateur et à sa perception de la qualité du produit.

L'information constitue un préalable pour tout usage fructueux d'une base de données. À ce sujet, il faut saluer le souci de transparence et de clarté de ProQuest : outre une liste de domaines couverts, une description générale présente le type de données indexées ainsi que les outils proposés, comme les thésaurus ou l'indexation bilingue. La fréquence des mises à jour est également explicitée. Ebsco et ProQuest ont développé tout un éventail d'outils de formation qui concernent, toutefois, presque toujours la plateforme en général et ne sont donc pas toujours très adaptés aux spécificités plus raffinées des différents produits. Par contre, les manuels s'avèrent assez développés, chez les deux sociétés, mais certaines ressources d'instruction sont davantage destinées aux professionnels de l'information qu'à l'utilisateur lambda. C'est pourquoi les bibliothèques de l'ULg offrent divers types de formations, dont le contenu s'adapte en fonction des besoins et du degré de spécialisation du public cible. ProQuest témoigne cependant d'un grand souci de didactique : l'interface de recherche simplifiée s'accompagne de liens vers les FAQ et vers un formulaire permettant de contacter un spécialiste pour des questions plus complexes. L'aide, stratégiquement placée juste au-dessus du formulaire de recherche, reste constamment accessible. ProQuest a également développé des didacticiels sur *YouTube*, que

l'Université de Liège a décidé de mettre en lien direct avec l'interface.

La recherche se trouve, logiquement, au centre des deux plateformes. Les menus en haut de l'écran permettent, d'une part de naviguer entre les différentes modalités de recherche et, d'autre part, d'accéder à des fonctionnalités plus personnalisées, liées à l'affichage et au compte d'utilisateur personnel.

La différence essentielle entre les deux outils réside dans l'association des données. Dans l'interface d'Ebsco, les différentes fonctionnalités, comme les index et le thésaurus, sont proposées séparément, tandis que ProQuest intègre tous ces outils dans l'interface de recherche. L'expérience de recherche se fait ainsi beaucoup plus globale. L'interaction entre les différents types de données étant beaucoup plus direct, l'impact est sensible au niveau de toutes les modalités de recherche. La recherche syntaxique, par exemple, donne une réelle impression de fluidité.

Les interfaces, tant chez Ebsco que ProQuest, sont conçues pour permettre un passage aisé d'une base de données à une autre. En haut de l'écran de recherche de la plateforme Ebsco, l'utilisateur peut choisir les bases de données qui seront sujettes à recherche. Ebsco a développé l'outil *EBS-COhost Integrated Search*, que la bibliothèque peut choisir d'utiliser. Chez ProQuest, une simple fenêtre permet ainsi de sélectionner une seule ou plusieurs des bases de données de la plateforme à interroger simultanément. Par contre, la fédération peut atteindre la précision de la recherche. En effet, certaines fonctionnalités sont propres à une base de données. Dans l'interface de ProQuest, un indice à l'intérieur de l'écran de recherche informe l'internaute que certains critères ne s'appliquent pas à toutes les bases. Chez Ebsco, le système supprime les critères non partagés par l'ensemble des bases de données utilisées. Tel est notamment le cas des thésaurus. Ceux-ci étant très spécifiques, ils peuvent limiter la recherche, même lorsque l'utilisateur a sélectionné plusieurs bases. En effet, le thésaurus n'interrogera que les bases particulières auxquel-

les ses mots-clés s'appliquent. Il importe de sensibiliser les utilisateurs à cette particularité. ProQuest fournit un effort en la matière par un message d'avertissement très explicite. Dans le cas d'Ebsco, le système propose les différents thésaurus des bases sélectionnées, en indiquant à quelle base ils appartiennent. De même, la fonctionnalité "names as subjects" de *MLA* reste disponible quand plusieurs bases sont interrogées en même temps, mais la recherche ne portera que sur *MLA*.

La lisibilité des résultats est très développée chez Ebsco. Les icônes identifiant les différents types de documents ou proposant des fonctionnalités complémentaires sont sobres et efficaces. Par ailleurs, à partir de la liste des résultats, l'internaute bénéficie d'un accès immédiat à toutes les informations permettant d'évaluer la pertinence des différentes références. À cet effet, l'affichage des sujets est particulièrement utile. ProQuest, par contre, propose une fonction d'aperçu, qui permet de visualiser immédiatement toutes les informations nécessaires. Les notices détaillées sont à la fois compactes et compréhensibles dans l'interface d'Ebsco. Chez ProQuest, la description peut s'étendre sur plusieurs pages. Par contre, le petit histogramme de la chronologie des résultats proposé par Pro-

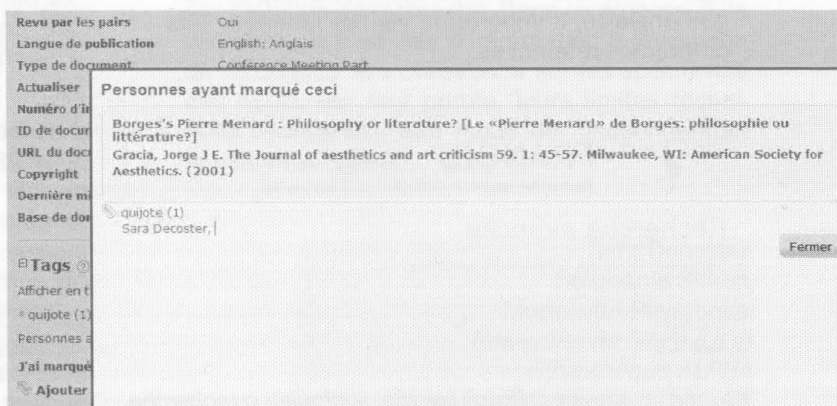


Fig. 4 : Les tags peuvent être partagés et s'intègrent dans la notice.

Quest ajoute un réel surplus d'efficacité. Du reste, Ebsco permet de formater l'affichage de la liste des résultats via la fonction "page option". Chez ProQuest, c'est en se connectant à son compte personnel que l'utilisateur modifie ses préférences. L'internaute est obligé de changer les préférences d'affichage sur l'interface de ProQuest. D'une manière générale, Ebsco propose de nombreuses possibilités pour modifier la présentation de l'écran de recherche. Chez ProQuest, les préférences de l'utilisateur sont gérées comme un ensemble à partir du compte personnel.

Mes tags (1)

1. quijote (1)
Premier ajouté : avril 09 2013
- Ne pas partager Modifier Supprimer

Fig. 5 : L'utilisateur peut gérer ses tags par son compte personnel.

La langue de l'interface est paramétrable pour toutes les bases de données. La traduction française pourrait cependant être plus soignée dans les bases de données accessibles par Ebsco. Si différents critères de recherche continuent à apparaître en anglais, il est surtout déroutant que les problèmes linguistiques affectent la compréhension de l'écran de recherche. Lorsqu'une variable passe de "off" à "on" dans les options préférentielles proposées en anglais, il passe au statut "le" dans la version française... D'ailleurs, à l'exception de *Francis*, les différentes bases de données fonctionnent principalement en anglais, même si *LLBA* propose un thésaurus en français et en espagnol. *Francis* propose des mots-clés en anglais et en français, pour toutes les notices. Une minorité de notices possède également une indexation matière en espagnol. L'ensemble des descripteurs est disponible dans une seule et même liste ; l'équivalence exacte des traductions entre les langues n'étant pas assurée, c'est à l'utilisateur de choisir les différentes occurrences de son terme de recherche. Dans une même langue, les sujets ne sont pas toujours utilisés d'une manière univoque. La plupart des résumés fournis par *LLBA*, par contre, sont en anglais. Le système propose une traduction automatique en français, qui se révèle assez déficiente. Comme la base de données transpose également tous les titres des références en langue anglaise, il importe de vérifier la langue des différentes références.

Le système se veut propice au partage sur les réseaux sociaux. Or, l'interactivité s'est implantée plus profondément dans le système. Les chercheurs peuvent ajouter des tags aux notices et éventuellement les communiquer à d'autres scientifiques actifs dans les bases de données ProQuest. Chacun peut créer ses tags personnels, mais l'échange avec toute la communauté passe par la création d'un profil public, dans lequel apparaissent notamment l'affiliation et l'adresse e-mail de l'utilisateur. En d'autres termes, la pleine exploitation des tags, visibles pour les utilisateurs du monde entier, doit se faire dans la transparence. L'application d'un tag public passe par un lien, permettant également la création de listes partagées. De cette manière, les différents utilisateurs de la base peuvent observer les activités des autres. Les tags relèvent d'une réelle interaction scientifique et sont susceptibles de favoriser les contacts entre chercheurs. Plus qu'un simple outil, la base de données devient réseau social.

En guise de conclusion : les outils de recherche et la bibliothèque de l'avenir

La différence entre les deux plateformes montre les enjeux de l'évolution des technologies de l'information.

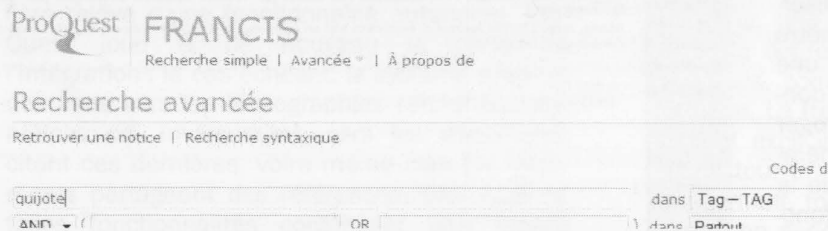


Fig. 6 : Les tags sont sujets à recherche.

L'exploitation des résultats s'avère variée sur les deux plateformes. Les deux plateformes proposent l'impression, l'envoi par mail et la citation, éventuellement à l'aide du logiciel *EndNote*. La technologie OpenURL établit le lien vers d'autres outils. Les ressources électroniques peuvent, de cette manière, être téléchargées directement. La même technologie permet la connexion directe vers le catalogue et le répertoire institutionnels. Bien sûr, les utilisateurs peuvent également enregistrer des notices ou des équations de recherche sur leur compte personnel. Ainsi, il devient possible de créer des alertes ou des flux RSS.

À ce sujet, ProQuest semble avoir adopté la philosophie 2.0. En premier lieu, l'interface propre à l'utilisateur est particulièrement élaborée et le

Sans tomber dans la caricature, ProQuest met l'accent sur une recherche à la fois intégrée et personnalisée. La plateforme propose d'office une expérience de recherche à la fois globale et interactive. Par ailleurs, les index de *Francis* et *LLBA* restent propres à la base de données, ce qui évite la surcharge

causée par des informations non pertinentes. Les thésaurus restent liés aux outils pour lesquels ils ont été conçus. Une information concernant la portée limitée du thésaurus s'insère directement dans l'écran de recherche. L'interactivité présuppose des questions semblables, illustrées très clairement par le cas de la saisie semi-automatique, que l'internaute peut, au demeurant, désactiver. La saisie semi-automatique n'est pas liée à l'indexation et n'est pas non plus déterminée par la base interrogée : elle est conditionnée par les recherches les plus fréquentes formulées par l'ensemble des utilisateurs de ProQuest. Les explications à ce sujet sont présentes dans le manuel, mais ne sont pas visibles sur le site.

Les nouvelles technologies, centrées sur l'interactivité, posent d'une nouvelle manière les anciennes questions relatives à l'interprétation des données. L'évaluation de l'information implique une conscientisation par rapport à ses origines. Il s'agit d'un problème aussi vieux que la recherche : toute recherche est en partie déterminée par les outils utilisés. Les difficultés ne sont pas uniquement générées par des contraintes techniques, mais concernent également la couverture des ressources. En effet, chaque outil, aussi évolué qu'il soit, ne reste jamais qu'un outil, avec ses qualités et ses limites : aucun outil ne permet de répondre à toutes les questions. Quelle que soit la qualité des ressources, une recherche n'aboutit réellement que grâce au sens critique de l'utilisateur. C'est l'internaute qui doit diriger la recherche dans la base de données, et non la base de données qui doit diriger la recherche de l'internaute. Par exemple, les facettes sont précieuses pour affiner une recherche, mais ne permettent pas de faire l'économie de la récolte préalable des informations. Si les bases de données incluant des articles en "full text", telles que *CMMC*, sont extrêmement commodés, la présence du texte complet n'est pas prédictive de l'utilité du contenu. D'une manière générale, la recherche implique toujours une réflexion sur la qualité de l'information. D'où l'importance des indications concernant le "peer review" que les différentes bases s'efforcent de spécifier.

Nos analyses ont clairement mis en évidence la circonscription, floue par ailleurs, des différents produits. Si l'exhaustivité totale est une illusion, sans doute incompatible avec l'augmentation exponentielle de la masse de données disponibles, le périmètre de recherche est marqué par des critères géographiques ou linguistiques. D'autres facteurs sont encore plus aléatoires, les revues n'étant souvent indexées que partiellement. Si la couverture des données est toujours réductrice, par la force des choses, ce constat permet également de cerner la spécificité des différentes ressources et d'évaluer leur complémentarité. Celle-ci est réelle, mais la recherche totale est plus que jamais une utopie, dans un monde qui voit paraître un livre toutes les trente secondes²². Dans un effort de dénombrer la quantité d'informations disponibles à l'échelle

planétaire, un article publié dans *Science* en 2011 estimait la capacité d'enregistrement globale à 2.9×10^{20} octets en 2007²³. Dans un tel contexte d'explosion de l'information, l'intégration de différents outils de recherche ouvre des perspectives d'avenir.

Face au foisonnement de publications scientifiques, le chercheur est souvent obligé de multiplier les recherches afin de repérer une information précise. La recherche intégrée permet de résoudre au moins partiellement ce problème et équivaut donc à un gain de temps et un confort plus élevé pour l'utilisateur. Pour répondre à ce nouveau besoin, Ebsco et ProQuest ont respectivement développé les produits "Discovery" et "Summon". Ces nouvelles interfaces interrogent simultanément tout un éventail de ressources différentes. Une seule opération de recherche mène à un nombre de résultats significativement accru. Les outils *Discovery* conjuguent les caractéristiques des catalogues et des bibliographies, joignant aux notices des livres présents à la bibliothèque des références à des articles cités dans une multiplicité de sources différentes, pas nécessairement accessibles en "full text". L'internaute de demain devra donc, comme celui d'aujourd'hui, comme celui d'hier, réfléchir à sa propre démarche et poser un regard critique sur ses résultats de recherche. Dans ce processus, la bibliothèque a un rôle d'information à jouer, afin de renseigner ses utilisateurs sur les spécificités des outils, sur leur portée, leurs limites techniques éventuelles, leur mode de fonctionnement le plus avantageux.

Sara Decoster
Stéphanie Simon
Muriel van Ruymbeke
Université de Liège (ULg)
Bibliothèque ALPHA
Place Cockerill, 1, bât. A3
4000 Liège
sara.decoaster@ulg.ac.be
stephanie.simon@ulg.ac.be
mvanruymbeke@ulg.ac.be
<http://www.libnet.ulg.ac.be/fr/libraries/alpha>

Octobre 2012

Notes

- 1 Cet article est le résultat d'un travail d'analyse effectué en septembre 2012.
- 2 Plus d'informations sur cet outil peuvent être trouvées à l'adresse suivante : <<http://www.mla.org/bibliography>>

- 3 Il existe un article antérieur retraçant l'évolution de la version imprimée jusqu'à la version informatisée, qui remet en évidence un certain nombre de points techniques, cf. Tonkin, Humphrey. Navigating and Expanding the MLA International Bibliography. *Journal of Scholarly Publishing*, 2010, vol. 41, n° 3, p. 340-353.
- 4 Cela concerne seulement 21% de la liste des périodiques.
- 5 Il s'agit d'un vieux problème : en 1998, Péter Jascó a déjà fait le même constat : Jascó, Péter. Analyzing the journal coverage of abstracting/indexing databases at variable aggregate and analytic levels. *Library & Information Science Research*, 1998, vol. 20, n° 2, p. 133-151.
- 6 La liste des titres couverts, précisant les années d'indexation et de mise à disposition d'articles en texte intégral, est disponible à l'adresse suivante : <<http://www.ebscohost.com/academic/communication-mass-media-complete>>
- 7 Soit, à partir de 2009 au minimum.
- 8 La raison de cette cessation de l'indexation n'est jamais détaillée par les fournisseurs qui rendent par ailleurs ces limites peu visibles pour les utilisateurs.
- 9 La base de données *Francis* nous a posé problème sur cet aspect. En effet, la liste des titres dépouillés dans *Francis* n'a pu être trouvée que difficilement et ne comprend aucune information sur les dates de fin d'indexation. Pour cette base, nous n'avons donc pu nous baser que sur les dates de parution des revues.
- 10 102 titres ont été analysés, correspondant à l'ensemble des titres dont l'indexation a été suspendue et pour lesquels nous sommes parvenues à retrouver suffisamment d'informations concernant leur état de parution.
- 11 Échantillon de 44 titres à indexation suspendue. La taille restreinte de l'échantillon s'explique de par la difficulté pour l'utilisateur d'arriver à ces données : il faut, pour chaque titre, accéder à la notice détaillée dans le *MLA Directory of Periodicals* afin de vérifier l'état d'indexation et les dates de parution.
- 12 Ebsco et ProQuest ne peuvent s'engager sur des données dont ils ne sont que les fournisseurs d'accès et non les éditeurs, comme c'est le cas pour trois des bases analysées : *LLBA*, *Francis* et *MLA*.
- 13 Le public de la Bibliothèque de Philosophie et Lettres de l'Université de Liège est majoritairement francophone.
- 14 L'échantillon était composé d'un nombre de titres non dédoublonnés égal à 1119, correspondant aux périodiques au titre commençant par A dans les quatre bases.
- 15 Pour rappel : 580.486 notices dans *CMMC* au 25 juillet 2012.
- 16 Il existe un article datant de 2007, concernant l'histoire sociale de l'écrivain Borges, qui est décrit sans que le nom de ce dernier ne soit mentionné dans la zone des sujets. En l'occurrence, l'exemple provient de *Francis*, mais ce type de problème est susceptible de se présenter dans n'importe quelle base. L'importance d'une bonne indexation par sujets est également mise en évidence dans l'article : DeLong, Linwood; Su, Di. Subscribing to databases: How important is depth and quality of indexing? *Acquisitions Librarian*, 2007, vol. 19: n° 37/38, p. 99-106.
- 17 Tonkin, Humphrey. Navigating and Expanding the MLA International Bibliography. *Journal of Scholarly Publishing*, 2010, vol. 41, n° 3, p. 346.
- 18 Le critère est proposé dans *Francis*. Dans *LLBA*, le champ est présent dans les fiches descriptives. Dans ce dernier cas, il est légitime de supposer que la différence s'explique par la distinction appliquée sur l'ancienne plateforme CSA entre le vocabulaire contrôlé du thésaurus et les simples mots-clés. Par contre, *Francis* ne dispose pas d'un véritable thésaurus. Les descripteurs introduits dans la zone "identificateur / mot-clé" de cette base sont souvent très génériques.
- 19 Des remarques semblables ont été formulées par Tonkin, Humphrey. Navigating and Expanding the MLA International Bibliography. *Journal of Scholarly Publishing*, 2010, vol. 41, n° 3, p. 340-353. Cet article permet également de comprendre les choix historiques effectués dans la base de données *MLA*.
- 20 Cf. Tab. 5.
- 21 Boucher, Amélie. *Ergonomie web : pour des sites web efficaces*. 3^e éd. Eyrolles, 2011, p. 12.
- 22 Darnton, Robert. Old Books and E-books. *European Review*, 2007, vol 15, n° 2, p. 165-170.
- 23 Hilbert, Martin ; López, Priscila. The World's technological capacity to store, communicate, and compute information. *Science*, 2011, vol. 332, n° 6025, p. 60-65. DOI: 10.1126/science.1200970.