BibliMed : une approche 2.0 de la recherche documentaire biomédicale

François Boutin, enseignant à l'Université Montpellier 1, concepteur et réalisateur de BibliMed - françois.boutin@univ-montp1.fr

Introduction

L'ouverture au grand public de la base de données bibliographique Medline à travers l'interface PubMed a radicalement changé les habitudes de recherche documentaire biomédicale. Autrefois l'apanage des bibliothécaires, la recherche documentaire est maintenant à portée de tous. Les services de documentation des institutions donnent même la possibilité à leurs usagers de bénéficier à distance des abonnements de l'établissement de sorte à ce que chacun puisse poursuivre sa recherche à domicile en synchronisant si besoin sa bibliographie avec son outil de gestion bibliographique favori.

En parallèle l'avènement du web 2.0 a profondément transformé les comportements sur Internet et la manière de concevoir la recherche d'information. Nombreux chercheurs allant jusqu'à utiliser leur moteur de recherche généraliste préféré pour effectuer leur recherche scientifique.

Avec la multiplication des interfaces de recherche documentaire et l'ouverture des bases de données scientifiques tant publiques que privées, la question n'est plus de savoir si l'on va parvenir à trouver des articles répondant à sa problématique mais où et comment on va pouvoir récupérer rapidement une information pertinente.

Nous allons analyser dans cet article comment les non spécialistes de la documentation appréhendent l'interface « classique » de PubMed et en quoi l'offre d'interfaces alternatives 2.0 peut leur faciliter grandement la recherche d'information.

PubMed versus Google Scholar

Parmi les interfaces « gratuites » d'accès aux références de Medline les plus utilisées, on en trouve deux proposant des approches radicalement différentes : PubMed et Google Scholar.

PubMed et le thesaurus MeSH

Avec plus d'un milliard de requêtes annuelles, PubMed [1] offre une recherche structurée, de qualité, dans la base de données Medline. Chacune des 20 millions de références d'articles de la base de données bibliographique américaine est indexée par une dizaine de descripteurs issus du thesaurus MeSH (Medical Subject Heading) comprenant lui-même 26 000 termes.

Effectuer une recherche par descripteur plutôt qu'en texte libre permet d'optimiser la recherche : Ainsi si l'on cherche des articles traitant du cancer des os, on préférera l'expression officielle « Bone Neoplasms [MeSH] » plutôt que « bone cancer » que l'on pourrait être tenté d'utiliser. En effet, en effectuant une telle recherche, on risque de passer à côté d'articles intéressants indexés par l'expression MeSH mais ne mentionnant pas l'expression exacte « bone cancer ».

Medline propose non seulement une indexation des articles par des descripteurs MeSH mais leur associe également parfois des qualificatifs ou subheadings (par exemple : therapeutic use, diagnosis, complications...). Au nombre de 83 ils permettent d'affiner la requête en privilégiant un point de vue. Si l'on recherche des articles sur les traitements du cancer des os on pourra utiliser la requête suivante : « Bone Neoplasms / drug therapy [MeSH] ».

L'interface PubMed propose également une panoplie de filtres facilitant la recherche d'information, notamment, entre autres, des filtres par auteur, date de publication, revue, ou type de publication. La combinaison de ces différents termes et filtres se fait par les opérateurs booléens : AND, OR, NOT.

Google Scholar: adaptation d'un moteur généraliste

Google Scholar [2] n'est pas à proprement parler une interface de recherche de Medline, toutefois nombreuses de ses requêtes aboutissent à des références d'articles mises à disposition par PubMed. Ainsi de manière très simple, « à la Google », l'utilisateur peut accéder à un contenu scientifique.

Contrairement à PubMed, Google Scholar offre très peu d'options de recherche: L'utilisateur saisit sa requête en langage naturel et le moteur de recherche propose une liste de résultats ordonnés de manière plus ou moins « obscure » (Google ne publiant pas ses recettes). Quand on sait que la majorité des consultations se limite à la première page de résultats, on est en droit de s'interroger sur la pertinence du classement proposé.

Contrairement à PubMed qui propose des références d'articles indexés par des spécialistes de l'indexation biomédicale, Google Scholar offre une indexation automatique des documents sans s'appuyer sur un thesaurus de référence. Par ailleurs, alors que PubMed référence des articles de qualité scientifique reconnue (avec comité de lecture), Google Scholar propose des sources plus larges d'information, ce qui pose le problème de la qualité scientifique des résultats.

En revanche, pour chaque article, Google Scholar précise combien et quelles références citent l'article en question, ce que ne permet pas PubMed.

PubMed ou Google Scholar?

Le choix semble s'imposer pour une recherche documentaire scientifique de qualité. Et pourtant, de nombreux chercheurs se limitent à l'utilisation de Google Scholar et en sont très satisfaits. D'autres utilisent PubMed « à la Google » en ignorant l'existence du thesaurus MeSH. D'après une étude menée sur 2,7 millions de requêtes durant 24 heures, à peine plus d'un pourcent des requêtes contiennent des termes du thesaurus [3]. Ca laisse entrevoir la pauvreté des requêtes menées par les non spécialistes.

Pourquoi un tel gâchis ? Pour quelle raison ce travail titanesque d'indexation (2 000 articles par jour) est-il à ce point sous exploité ? Existe-t-il des alternatives crédibles à ces deux interfaces ? Quel intérêt ont les chercheurs à les utiliser ? Vous trouverez des éléments de réponse intéressants dans l'excellent blog d'Hervé Basset [4]. C'est aussi l'objectif de cet article d'essayer de répondre à ces questions.

Interfaces alternatives « classiques »

Depuis quelques années, PubMed offre non seulement une interface web pour les chercheurs mais également une interface de programmation (API) à l'attention des programmeurs leur permettant d'accéder librement à la base de données de Medline et de construire leur propre interface de Medline. Une liste d'interfaces alternatives est ainsi accessible sur le site de PubMed [5].

HubMed: compromis entre richesse et simplicité

Parmi ces interfaces, nombreuses ont choisi un compromis entre la richesse de PubMed et la simplicité de Google Scholar. HubMed [6] par exemple est une interface 1.0 sobre et simple d'utilisation proposant des fonctionnalités de PubMed : sauvegarde, export, articles en ligne (full text), ou associés (related)...

Si ces interfaces ont le mérite d'être simples à utiliser, elles n'apportent pas réellement de valeur ajoutée ce qui explique certainement leur manque de popularité.

PubGet: une interface 2.0 répondant au besoin immédiat d'information

PubGet [7] est l'une des alternatives 2.0 les plus populaires de PubMed : elle revendique un million d'utilisateur et 400 institutions. Son principe est simple : donner un accès immédiat aux articles en téléchargement (suivant les droits d'accès de l'établissement). Cet outil répond au besoin « impulsif » des usagers 2.0 d'une information directement utilisable.

Cette interface intègre des outils du web 2.0 (ajout de commentaires, tweets, flux RSS) mais ne permet pas de recherche avancée (avec termes MeSH, filtres,...) ce qui fait défaut à un chercheur rigoureux.

En effet, contrairement aux solutions décrites ci-dessous, PubGet ne propose pas dans son interface de fonctionnalités sémantiques pouvant faciliter la recherche (telles que des suggestions contextuelles).

Interfaces alternatives « sémantiques »

Doit-on nécessairement accepter un compromis entre richesse d'interface et simplicité d'utilisation ou peut-on concevoir des interfaces alternatives qui proposent tout à la fois richesse et simplicité ?

Pour dépasser ce compromis, certaines solutions alternatives mettent à profit à la fois les bonnes pratiques du web 2.0 mais aussi des fonctionnalités sémantiques.

Les interfaces sémantiques : richesse et simplicité d'utilisation

Les interfaces alternatives sémantiques cherchent à exploiter la richesse des fonctionnalités de PubMed tout en évitant sa « lourdeur » d'utilisation [8]. Pour cela, elles mettent à profit les métadonnées associées aux résultats de la recherche telles que descripteurs MeSH, auteurs, revues...

Alors qu'avec PubMed, il faut naviguer dans l'arborescence du thesaurus MeSH pour construire des requêtes de qualité, il suffit avec les interfaces sémantiques d'explorer les termes contextuels proposés. C'est l'une des grandes forces de ces interfaces de faire ressortir notamment les concepts clés du domaine (mais aussi les auteurs et journaux principaux) qui permettront d'itérer la recherche.

Ainsi, les interfaces « sémantiques » proposent un contenu enrichi favorisant la construction de savoir par l'élaboration d'un réseau de méta connaissances. Par exemple, en cherchant des documents sur « Bone Neoplasms », on se voit proposer les termes clés associés (comme « Osteosarcoma ») mais aussi la revue la plus pertinente (« Cancer »).

On parle d'interfaces « sémantiques » car elles permettent de construire du « sens » et pas uniquement de fournir une liste de résultats comme peuvent le faire les interfaces « classiques ».

Quertle: relations entre concepts

Quertle [9] est une interface de recherche soignée très simple d'utilisation qui propose une liste de termes clés cliquables (« key concepts ») et différents filtres. Sa particularité réside dans le fait qu'elle propose une liste de résultats sémantiques (onglet « relationships ») associant dans une même phrase les différents termes choisis ou éventuellement des synonymes.

Cependant, cette interface présente des faiblesses. Notamment, les termes suggérés ne proviennent pas nécessairement du thesaurus MeSH. Par ailleurs, Quertle ne permet pas une recherche itérative : en effet, après avoir cliqué sur un concept clé suggéré, on ne peut pas ajouter un nouveau concept sans enlever le précédent. Ce qui est problématique si on veut mener une recherche efficace.

NextBio: recherche itérative

L'interface NextBio [10] donne une importance toute particulière aux métadonnées dans son interface. L'utilisateur a le choix entre différentes catégories (descripteurs MeSH, tissus, substances, gènes et protéines, types de publication, organismes...).

Après avoir choisi une catégorie, l'utilisateur se voit suggérer une liste de termes associés (sous forme de nuage de tags). Par exemple pour la requête « bone cancer », si on choisit la catégorie « gene and proteins », les tags les plus importants proposés sont « P2RX7 » et « TNFSF11 ».

Cette interface n'offre malheureusement pas une recherche par qualificatif. Ainsi, on ne peut préciser simplement si l'on s'intéresse à l'utilisation thérapeutique d'un traitement ou à ces effets indésirables.

GoPubMed: interface de recherche pour experts et veilleurs

A travers une interface très recherchée, GoPubMed [11] met à disposition différents outils pour mener une recherche documentaire de qualité. Il donne notamment accès à une ontologie « dynamique » dénombrant pour chaque terme de l'arborescence le nombre d'articles traitant à la fois du terme en question et de l'équation de recherche actuelle. On apprend ainsi que sur les 100 000 articles traitant de « Bone Neoplasms [MeSH] », 50 articles traitent également de « Intestinal Mucosa ». Notez que pour obtenir ce dernier terme, on a dû parcourir l'arborescence : Anatomy / All of Anatomy / Tissues / Membranes / Mucous Membrane / Intestinal Mucosa. Si le tour de force technologique est indéniable, l'exploration fastidieuse de l'arborescence laisse douter de son utilité pour des recherches classiques.

En revanche, l'outil est incontestablement très intéressant pour des recherches d'experts. D'autant qu'il permet également d'accéder à des informations statistiques telles que la liste des journaux, auteurs et pays les plus en vogue dans le domaine. GoPubMed propose aussi une cartographie du réseau de co-auteurs, à condition que le nombre de résultats ne soit pas trop important. Si cette représentation peut présenter un intérêt à premier abord, elle donne beaucoup trop d'importance aux auteurs ayant été nombreux à participer à un même article (ce qui n'est pas nécessairement judicieux). GoPubMed tente de corriger ce problème en donnant un poids plus faible à ces liens (en jouant sur leur couleur).

Tout comme NextBio et Quertle, GoPubMed ne permet pas de tirer partie des qualificatifs (subheadings) de Medline ce qui empêche de répondre simplement aux questions : « Quel traitement ? Quel diagnostic ? Quelles complications ? Quel effet psychologique ? Quelles contre indications ?... ».

BibliMed : une interface sémantique simple et efficace

Pour avoir formé durant dix ans des étudiants en fin d'étude odontologiques aux techniques de recherche documentaire et plus particulièrement à l'utilisation de PubMed, l'auteur s'est rendu compte de la nécessité de proposer des interfaces alternatives à la fois plus accessibles et plus efficaces.

BibliMed [12] est une interface développée depuis 2010 par l'auteur reposant sur des techniques 2.0 et sémantiques. Elle cherche à favoriser des recherches de qualité de manière intuitive [8]. Testée auprès d'étudiants elle a donné des résultats très encourageants qui seront détaillés dans ce chapitre.

Pourquoi une nouvelle interface alternative?

Les interfaces alternatives mentionnées précédemment facilitent la recherche documentaire pour les non spécialistes de la documentation. Pourtant, elles ne répondent pas à toutes les questions. BibliMed propose d'intégrer des fonctionnalités originales répondant aux besoins des utilisateurs :

- Aide à la formulation des requêtes en français ou en anglais
- Prise en compte des qualificatifs (pas uniquement des descripteurs MeSH)
- Gestion simplifiée de l'historique des recherches et de l'export
- Liens vers des contenus contextuels

Interface en français

Une grande difficulté que rencontrent les utilisateurs français est liée à la construction de la requête en langue anglaise. Pour peu que l'usager fasse une recherche en texte libre avec une expression approximative (traduite mot pour mot du français), il risque de ne pas obtenir une réponse satisfaisante.

BibliMed évite ce problème en suggérant, lors de la saisie, des descripteurs MeSH ou leurs synonymes en français ou en anglais. Il suggère également des substances (et leur descripteur MeSH associé) ainsi que des titres de revues (voir Figure 1 les suggestion associées à « Carie de... »).



Figure 1 - Suggestion de termes MeSH, substances et revues à la saisie

La page de résultats est accompagnée d'un nuage de mot clés (Figure 2). Pour chacun d'eux, BibliMed offre la traduction française officielle de l'INSERM et la définition de WikiPedia en français et en anglais.

Prise en compte des qualificatifs Medline

BibliMed est, à ce jour, la seule alternative de PubMed offrant la gestion des qualificatifs (subheadings). Notons que chaque article est référencé par des descripteurs pouvant être associés à des qualificatifs.

L'utilisation judicieuse de ces qualificatifs par BibliMed permet d'optimiser les recherches. Par exemple « Bone Neoplasms / diagnosis [MeSH]» renverra des articles traitant du diagnostic du cancer des os alors que « Bone Neoplasms / etiology [MeSH] » s'intéressera à l'étiologie de la maladie.

BibliMed propose une utilisation originale des qualificatifs en suggérant pour un qualificatif choisi tous les descripteurs associés. Par exemple, si la requête contient l'expression « Dental caries [MeSH] », en choisissant le qualificatif « therapeutic use », il nous est suggéré les descripteurs : « Cariostatic Agents », « Fluorides », « Toothpastes », … pour itérer la recherche (voir Figure 2).



Figure 2 - Utilisation thérapeutique des fluorures contre la carie dentaire

BibliMed offre aussi une fonctionnalité non proposée par PubMed permettant d'associer un qualificatif à une substance. Ainsi on peut chercher par exemple les propriétés chimiques (« chemistry ») ou thérapeutiques (« therapeutic use ») des bio-os (« bio-oss : [substance] »).

Gestion simplifiée de l'historique et de l'export

BibliMed simplifie à l'extrême l'enregistrement des références bibliographiques : l'utilisateur n'a qu'à cliquer sur l'icône d'enregistrement à côté de l'article souhaité (voir Figure 2).

La référence est alors automatiquement enregistrée dans une rubrique ayant pour nom, celui de la requête. Rubriques et notices bibliographiques associées sont alors présentées sur la page d'accueil (voir Figure 3). Un simple clic sur le nom de la rubrique permet de relancer la requête.

Un clic sur le titre de la notice permet de la dérouler. Les notices peuvent ensuite être facilement exportées sous différents formats (Zotero, Mendeley, Medline, Reference Manager,...).



Figure 3 - Equations, articles et mots clés favoris

Liens vers des contenus contextuels

BibliMed donne accès non seulement aux notices Medline mais aussi à divers contenus contextuels (voir Figure 2). Notamment dans la partie droite de l'interface, BibliMed présente des livres en relation avec le dernier terme saisi ou sélectionné. L'usager pourra localiser les livres dans SUDOC (l'interface des bibliothèques universitaires françaises), dans WorldCat (Catalogue de bibliothèque mondial) ou encore les rechercher avec Amazon. Les références pourront être stockées dans Mendeley ou Zotero.

BibliMed propose également d'autres contenus contextuels au survol de chaque terme MeSH suggéré :

- Recherche simultanée (« Entrez ») dans toutes les bases de données de Medline (« PubMed »,
 « Nucleotid », « Protein », « Genome », « Structure », « Taxonomy », « PubChem substance »...)
- Le site d'essais cliniques : clinicaltrials.gov
- Des bases génomiques et pharmacologiques

BibliMed: retours d'expérience

Contexte

Deux promotions de 60 étudiants en fin d'études odontologiques ont suivi une formation de deux heures à la recherche documentaire incluant 1h30 de formation à PubMed, SUDOC et Zotero et 30 minutes de présentation de BibliMed. La première promotion a été formée il y a 18 mois, la seconde il y a 6 mois. Depuis, l'auteur n'a pas rencontré les étudiants de ces promotions, à l'exception de quelques contacts individuels. Ainsi les étudiants ont utilisé librement l'interface de recherche bibliographique de leur choix pour leur thèse (sans « contrainte » ou encouragement de l'auteur).

Analyse des fichiers logs (historique des événements du site web BibliMed)

L'analyse porte sur les 8 838 requêtes lancées par 109 étudiants ayant utilisé BibliMed « seuls » dans les semaines ou mois ayant suivi la formation bibliographique entre le 11/10/2010 et le 03/04/2012.

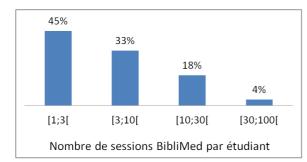
Précisons le vocabulaire utilisé : On parlera de nouvelle requête dès qu'il y a ajout ou suppression d'un terme ou d'un filtre. Notez que les requêtes menées durant la période de formation ont été écartées.

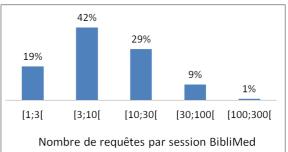
On dira qu'un étudiant a ouvert une session, dès lors qu'il s'est connecté sur le site et a effectué au moins une recherche dans la journée (on comptera au maximum une session par jour et par étudiant).

91% des étudiants ont utilisé BibliMed suite à leur formation. Les habitudes d'utilisation du site sont synthétisées dans les diagrammes en Figure 4. Nous présentons quelques résultats synthétiques :

- En moyenne 6,7 sessions par étudiant et 12,2 requêtes par session (81 requêtes par étudiant).
- 55% des étudiants ont eu 3 sessions ou plus. 22% des étudiants en ont eu au moins 10.
- 39% des sessions comprennent au moins 10 requêtes. 10% en comptent plus de 30.

Ainsi nombreux étudiants sont venus à diverses reprises sur le site de BibliMed et ont construit des requêtes itératives avancées (c'est ce que confirme une analyse qualitative de ces recherches).





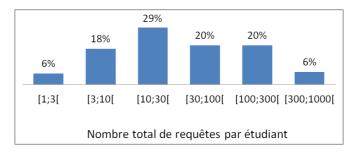


Figure 4 - Statistiques d'utilisation du site BibliMed par les étudiants

44% des termes ajoutés sont des termes MeSH (voir Figure 5). Parmi eux, 20% sont associés à des qualificatifs (les plus utilisés : methods, etiology, diagnosis, therapeutic use, anatomy & histology,...).

Ce résultat souligne la pertinence de l'interface BibliMed : valoriser l'utilisation des descripteurs et qualificatifs issus du thesaurus de Medline pour optimiser la qualité des recherches.

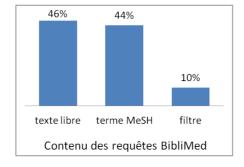


Figure 5 - utilisation significative des termes MeSH

Une étude qualitative menée (avec la deuxième promotion) a fait très nettement ressortir deux adjectifs : « simple et efficace » (qui accompagnent depuis le logo de BibliMed).

Conclusion

Si l'interface PubMed est généralement appréciée des spécialistes de la documentation, elle est souvent perçue comme rebutante par bon nombre de chercheurs et d'étudiants ne maîtrisant ni les subtilités de l'interface ni la langue de Shakespeare. Certains se replient sur des interfaces alternatives plus simples d'utilisation mais souvent beaucoup plus pauvres. D'autres persistent à utiliser PubMed « à la Google ». Il en résulte des requêtes fades donnant trop souvent des résultats peu satisfaisants.

Certaines solutions alternatives 2.0 cherchent à concilier simplicité d'utilisation et qualité de recherche en s'appuyant sur une interface sémantique tirant le meilleur parti des métadonnées.

Parmi ces solutions, BibliMed suggère des termes MesH et synonymes en français et anglais facilitant l'élaboration itérative de l'équation de recherche. Elle offre aussi divers filtres et outils contextuels pour un accès rapide à une information de qualité. Enfin, c'est la seule interface alternative qui propose une recherche par qualificatif permettant de répondre à des questions du type « Quels sont les effets adverses de tel traitement ? ».

Une analyse quantitative de milliers de requêtes d'étudiants en fin d'études odontologique a montré la pertinence de l'interface BibliMed pour un public de jeunes chercheurs et la richesse des équations construites pas à pas. Ces étudiants ont également souligné la simplicité et l'efficacité de BibliMed.

L'intérêt de BibliMed dépasse largement la simple recherche d'articles biomédicaux. L'enjeu est d'offrir, à travers une interface intuitive, un contenu enrichi pertinent, favorisant la construction itérative de méta-connaissance indispensable au scientifique de demain.

Bibliographie

- [1] « PubMed NCBI ». [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed.
- [2] « Google Scholar ». [Online]. Available: http://scholar.google.fr/.
- [3] J. R. Herskovic, L. Y. Tanaka, W. Hersh, et E. V. Bernstam, « A day in the life of PubMed: analysis of a typical day's query log », *J Am Med Inform Assoc*, vol. 14, n°. 2, p. 212–220, avr. 2007.
- [4] H. Basset, « Science Intelligence and InfoPros », *Science Intelligence and InfoPros*. [Online]. Available: http://scienceintelligence.wordpress.com/tag/pubmed.
- [5] Lu, Database 2011, « Biomedical Literature Search Tools », *NCBI, National Library of Medicine*. [Online]. Available: http://www.ncbi.nlm.nih.gov/CBBresearch/Lu/search/.
- [6] « HubMed ». [Online]. Available: http://www.hubmed.org.
- [7] « Pubget: the search engine for life-science PDFs ». [Online]. Available: http://pubget.com/.
- [8] H. Basset, D. Stuart, et D. Silber, From Science 2.0 to Pharma 3.0: Semantic Search and Social Media in the Pharmaceutical Industry and STM Publishing. Chandos Publishing (Oxford) Ltd, 2012.
- [9] « Quertle® Intelligent semantic queries of MEDLINE (PubMed) and the biomedical literature ». [Online]. Available: http://www.quertle.info.
- [10] « NextBio ». [Online]. Available: http://www.nextbio.com.
- [11] Transinsight, « GoPubMed® ». [Online]. Available: http://www.gopubmed.org.
- [12] « BibliMed ». [Online]. Available: http://biblimed.fr.