

## **COÛT ET COMPÉTENCE, L'EFFET MATTHIEU DANS LA RECHERCHE BIBLIOGRAPHIQUE EN LIGNE**

**Magdeleine MOUREAU**

L'interrogation des bases de données en ligne a fortement contribué à établir un lien entre la recherche d'une information et son prix.

L'utilisateur potentiel n'associe pas facilement la recherche d'une information et un prix à payer. **Dans la recherche de références en ligne on paie un moyen et non un résultat et il n'y a aucune corrélation entre le prix du moyen et la valeur du résultat.**

L'information trouvée (le résultat qui va être facturé) correspond à un des cinq cas suivants :

- 1) l'information est pertinente, nouvelle et accessible,**
- 2) l'information est pertinente mais déjà connue de l'utilisateur,**
- 3) l'information est pertinente mais inaccessible**  
soit : - intellectuellement  
soit : - physiquement
- 4) l'information est inintéressante et non pertinente,**
- 5) aucune information n'est retrouvée.**

Dans une recherche il faut donc, si l'on veut éviter de payer un service qui n'aurait pas atteint son but, se situer majoritairement dans le cas 1, éliminer si possible le cas 2, ne se situer dans le cas 3 qu'avec circonspection, éviter totalement le cas 4 et être sûr que le cas 5 n'est pas le résultat d'une recherche insuffisante. Face à la multiplication de documents, quels sont les moyens dont dispose l'utilisateur pour effectuer une sélection ? Il a deux approches sur lesquelles faire porter son effort l'approche stratégique et l'approche conceptuelle.

Chacune de ces approches suppose la mise en oeuvre d'une batterie de moyens.

### **L'approche stratégique**

Le premier acte d'une stratégie bien menée est la sélection de la base ou des bases de données à interroger. De ce choix initial dépend une bonne partie du rapport performance/coût de la recherche.

Dans la première moitié des années 70 l'utilisateur des banques et bases de données (BBD) en ligne disposait de quelques dizaines de base de données, au début des années 80 de plusieurs centaines au début des années 90, il faut maintenant compter en milliers. De nouvelles couches de connaissances se superposent donc sans cesse à celles qui sont déjà nécessaires et qu'il faut continuellement remettre à jour.

Un outil stratégique qui peut être fort efficace sera, quand il existe, **le fichier des fichiers** (DBI chez SDC, DIALINDEX chez DIALOG, CROSS chez BRS, QUESTINDEX chez ESA/IRS).

Le fichier des fichiers qui regroupe le vocabulaire présent dans l'ensemble de bases de données peut permettre de sélectionner un fichier auquel on n'aurait pas pensé a priori, il permet de sélectionner le ou les fichiers où les termes d'interrogation sont les plus employés et, donc, où il sera possible de trouver le plus grand nombre de réponses satisfaisantes.

Cependant cet outil n'est pas la panacée et peut conduire à bien des déboires. Un fichier ainsi sélectionné mais parfaitement inconnu tant dans son contenu que dans sa structure peut conduire à une recherche parfaitement infructueuse quoique coûteuse.

Le deuxième élément stratégique est **la bonne connaissance du logiciel d'interrogation**. Les modalités de procédures s'oublient très rapidement pour qui reste quelque temps sans pratiquer et les quelques secondes de réflexion passées à retrouver la commande exacte finissent par peser lourd sur le prix de revient d'une recherche si elles se multiplient. Seule une connaissance approfondie permet d'utiliser au mieux les commandes spécifiques du logiciel ; bien sûr les systèmes qui offrent des menus (comme Techdata sur BRS) seront une aide appréciable pour celui qui n'est pas très sûr de sa mémoire et de ses connaissances. Mais les systèmes à menus ajoutent à la recherche un surcoût qui peut être très important et multiplier le temps de un à quatre.

Le troisième élément sera la capacité de l'interrogateur à élaborer **une équation de recherche efficace**, par exemple le choix des "filtres" que l'on utilisera soit pour élargir, soit pour réduire le nombre des références obtenues. Ce pourra être : les mots-clés sélectionnés pour l'interrogation qui selon leur nombre d'occurrences seront élargis ou limités :

- les noms des auteurs ou d'organismes
- le nom ou le type de publication
- la langue
- la date
- la classification, les codes
- la pondération, quand elle existe.

Il s'impose donc à l'évidence que, si une approche stratégique bien menée peut conduire à de substantielles économies par rapport à une recherche qui n'utilise que les commandes les plus classiques et justifier plus que largement de son prix, elle ne peut être l'apanage que d'un utilisateur confirmé et ne pourra pas être développée par un utilisateur occasionnel, qu'il soit utilisateur intermédiaire ou utilisateur final.

**L'approche conceptuelle**, elle, est une d'une autre nature puisqu'il s'agit d'étudier le contenu sémantique de la question. Avec l'approche stratégique les éléments sont explicités, avec l'approche conceptuelle c'est l'implicite qui est en jeu.

Cette approche est nécessaire chaque fois que l'on s'aperçoit que les termes retenus pour la question n'apportent pas les éléments de réponse souhaités. Il s'agit de définir l'environnement sémantique d'un terme, avec ses relations de synonymie et de voisinage, et seul le spécialiste du sujet possède des connaissances qui lui permettrait de développer le thésaurus instantané qui lui permettra de mener à bien sa recherche, encore faudra-il que celui-ci corresponde avec le vocabulaire utilisé dans la base de données. Cependant pour le non spécialiste il existe au niveau des systèmes disponibles en ligne un certain nombre d'aides qu'il faut savoir exploiter. Certains sont conçus directement pour cet usage.

**Les thésaurus** développés pour les producteurs de bases de données peuvent pallier un certain manque de connaissances chez l'utilisateur. Matériellement un thésaurus présente une série de termes appartenant à un domaine précis et couvrant chacun un concept ou un ensemble de concepts déterminés, cette détermination opérant par le truchement des

relations de ce terme avec d'autres termes relation hiérarchique relation associative relation d'équivalence relations qui permettent d'en restituer la couverture sémantique.

Enfin des notes explicatives précisent l'emploi de certains descripteurs. Un thesaurus bien fait permet à un néophyte dans un domaine de formuler correctement une question dans ce même domaine.

**Les fichiers dictionnaires** en chimie sont un outil précieux pour un vocabulaire qui se caractérise par son grand nombre de synonymes et son ambiguïté. Ils sont complétés par les très puissants **systèmes de codage des composés chimiques**.

Il est en effet presque impossible de dégager un vocabulaire qui permettrait de sélectionner tous les membres d'une famille de composés issus par substitution d'une même molécule de base. La représentation canonique (c'est-à-dire unique) d'une structure donnée a débarrassé le chimiste d'un problème de nomenclatures. Les mêmes systèmes ont rendu possible la recherche de classes de substances ayant une même structure ou de mêmes combinaisons de structures. Ces systèmes sont désormais accessibles en ligne et couplés avec les systèmes de recherche bibliographique (DARC sur QUESTEL, CAS ONLINE sur STN). Ils ont introduit dans la recherche chimique une rigueur qu'il n'avait pas été possible d'atteindre jusqu'alors.

A ces moyens directs on peut ajouter une batterie de moyens indirects dont les plus importants sont :

L'utilisation des **co-occurrences entre termes** qui apporte une aide non négligeable à la reformulation des questions. Il est possible de demander d'associer aux mots-clés les plus significatifs les autres mots-clés qui les accompagnent le plus fréquemment, on peut demander la même information pour des noms d'auteurs et d'organismes. On peut ainsi préfigurer la démarche qu'adopterait un système expert (ZOOM sur ESA-IRS, GET sur INFOLINE, MEMSORT sur QUESTEL).

Par exemple une recherche sur les domaines dans lesquels était appliquée la géométrie fractale a permis de dégager très rapidement et facilement les termes les plus souvent utilisés avec "fractale". C'est ainsi que nous avons découvert le nombre important de co-occurrences entre **fractale** et **percolation**. Les documents ainsi indexés étaient ceux qui intéressaient les spécialistes en production des gisements, qui utilisant dans leur domaine le terme **drainage** pour les phénomènes de même

nature auraient spontanément associé **fractale et drainage** et n'auraient retrouvé aucun document.

**Les fichiers de citations** permettent aussi d'améliorer les recherches au niveau conceptuel. En effet, le vocabulaire utilisé dans des sciences ou techniques nouvelles peut évoluer très rapidement, certains termes devenant obsolètes et étant remplacés par des mots nouveaux que l'interrogateur ne connaît pas. C'est là un des moyens les plus sûrs de prendre connaissance d'un vocabulaire nouveau. Cependant les fichiers de citations demandent un préalable : avoir sur la question un document qui fasse autorité.

En effet, quand on connaît la référence d'un document qui couvre parfaitement le sujet sur lequel on recherche des documents, on peut inférer que les documents qui citeront ce document traiteront du même sujet que lui, tout en étant plus récents. L'étude de ces documents doit ensuite permettre de sélectionner le vocabulaire utilisé à cette date par les spécialistes du domaine, vocabulaire qui a pu se modifier avec le temps.

Enfin, c'est souvent le moyen le plus efficace de retrouver des références récentes sur un sujet dont on a du mal à isoler le contenu conceptuel et dans le domaine duquel on ne sait comment formuler les questions.

Dans l'approche conceptuelle, l'utilisateur final spécialiste du domaine de la question est le mieux armé au départ pour obtenir une belle performance. Mais elle peut être tout aussi bien le résultat de la recherche imaginative d'un utilisateur confirmé des bases de données.

### **L'effet Matthieu ou l'expert des systèmes**

Il est donc évident que les paramètres à intégrer pour mener à bien une recherche qui ait un bon rapport qualité-prix sont variés et variables. Seul le spécialiste en interrogation des BDD les connaît et les utilise à bon escient. L'utilisateur occasionnel ne peut, quelle que soit la connaissance du domaine qu'il interroge, atteindre les mêmes performances. C'est ce qui a donné naissance à un type de spécialiste qui se profilerait à l'horizon 1974, mais dont on n'était pas alors certain du devenir : l'expert en interrogation des banques de données en ligne (que ce soit à l'origine un intermédiaire ou un utilisateur final reconverti).

Une bonne recherche nécessite le plus souvent plusieurs bases de données. Seul l'intermédiaire professionnel qui a interrogé un nombre de

fichiers supérieurs (et parfois de beaucoup) à la centaine, peut devenir l'expert capable d'orienter sa recherche sur les fichiers les meilleurs et savoir comment les compléter par l'interrogation de fichiers moins évidents.

L'expert en interrogation réagit vite, maîtrisant le logiciel il comprend vite et il sait quoi faire. Il jongle avec les fichiers d'attente, les recherches croisées ou les stratégies sauvées et sera, dans le plus simple des cas, d'au moins 25 % plus rapide (pour une réponse souvent meilleure) que n'importe quel autre utilisateur. L'expert en base de données est par son expertise-même appelé à participer à des groupes de travail où s'élaborent et s'évaluent des bases de données et des logiciels. Il participe aux grands congrès documentaires, qui par leurs expositions et revues de produits sont l'endroit où il est facile d'acquérir en peu de temps une foule de connaissances sur des produits nouveaux qui peuvent se révéler miraculeusement utiles.

C'est ainsi que joue "l'effet Matthieu", d'après la parabole des talents, *Évangiles de St Matthieu 25/29* : "*Car on donnera à celui qui a et il sera dans l'abondance, mais à celui qui n'a pas, même ce qu'il a lui sera retiré.*" C'est-à-dire que le spécialiste qui fait de bonnes recherches est de plus en plus sollicité, il s'enrichit de plus en plus en expérience et devient encore meilleur. **Plus il en fait, mieux il le fait et plus on lui en demande.**

L'utilisateur occasionnel qui compare ou dont on compare les résultats en temps passé, résultats obtenus et factures à payer, délègue de plus en plus ses recherches et en finira par perdre toute activité dans ce domaine. C'est du moins l'évolution qui se dégage à l'Institut Français du Pétrole où la recherche en ligne a débuté en 1974. Avec le triple handicap d'une demande qui augmente en nombre et en complexité, de fichiers devenant sans cesse plus lourds et plus complexes et d'un budget fixe, c'était la seule voie ouverte pour maintenir le service demandé. Mais il faut pour cela une organisation suffisamment centralisée pour pouvoir contrôler les performances et les coûts.

Les quelques utilisateurs finals existants ont, au cours des années, abandonné l'interrogation des fichiers pour parfois ne conserver qu'un type d'interrogation bien spécialisé et dans un fichier bien précis, sur les structures chimiques par exemple. Les spécialistes en interrogations se sont encore plus spécialisés et sont devenus experts. En attendant l'hypothétique système expert capable d'orienter valablement et de conduire la recherche de l'utilisateur occasionnel, ne vaut-il pas mieux, si

la structure de l'organisation s'y prête, utiliser un expert des systèmes ? La qualité et le coût des recherches bibliographiques en seront les nets bénéficiaires.

Magdeleine MOUREAU

N.D.L.R. : Cet article, proposé à l'EPI, est paru sous une forme plus développée dans le *Bulletin des Bibliothèques de France*, tome 33, n° 3, 1988.