

Réflexions

Magdeleine Moureau

Directeur adjoint de la Direction Études et documentation
Institut français du pétrole

COÛT ET COMPÉTENCE

L'EFFET MATTHIEU DANS LA RECHERCHE BIBLIOGRAPHIQUE EN LIGNE*

AVEC la multiplication des OPACs se développent aujourd'hui des menus et des outils d'aide à l'interrogation de plus en plus conviviaux, qui permettent à tout utilisateur potentiel de dialoguer avec la base de son choix. Mais à quel prix ? Et pour quel résultat ? Nouveaux venus parmi les spécialistes de l'information, les experts des systèmes, spécialistes de l'interrogation des bases de données en ligne, tendent de plus en plus à s'imposer. Ils sont quatre à l'Institut français du pétrole. Porté très tôt vers les systèmes en ligne en raison de la grande fertilité de son domaine de recherche, l'IFP est à la pointe des recherches conceptuelles sur les outils linguistiques d'aide à l'interrogation et se distingue notamment par ses travaux sur les thésaurus.

L'interrogation des bases de données en ligne a fortement contribué à établir un lien entre la recherche d'une information et son prix. Cette notion de prix n'était jusqu'alors admise que lorsqu'il s'agissait de l'achat de produits papier (livres ou périodiques), dont la consultation pouvait d'ailleurs s'effectuer très souvent gratuitement dans les bibliothèques publiques. Alors que la notion de coût fut souvent associée à la notion d'information, celle de prix était toujours restée plus discrète : tout en acceptant intellectuellement l'idée qu'une information pouvait valoir très cher — les romans et les films d'espionnage avaient beaucoup fait pour vulgariser cette notion —, le sentiment qui prévalait généralement voulait que l'information soit gratuite.

Pour un meilleur rapport qualité/prix

L'utilisateur potentiel n'associe pas facilement la recherche d'une information et un prix à payer.

* Cet article est la mise à jour d'un article paru en anglais dans *Online review* sous le titre « Cost and know-how : the « Matthew effect » in information retrieval », 1987, vol. 11, n° 6.

Dans la recherche de références en ligne, c'est un moyen qui est facturé, non un résultat, et il n'y a aucune corrélation entre le prix du moyen et la valeur du résultat.

Exiger la qualité...

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

- l'information est pertinente, nouvelle et accessible ;
- l'information est pertinente, mais déjà connue de l'utilisateur ;
- l'information est pertinente, mais inaccessible pour l'utilisateur. Cette inaccessibilité peut avoir deux niveaux : elle peut être inutilisable intellectuellement, car d'un niveau qui suppose des connaissances scientifiques ou techniques supérieures à celles de l'utilisateur, ou écrite dans une langue indéchiffrable par l'utilisateur ; elle peut aussi être inaccessible physiquement : on ne sait ni où ni comment se la procurer.

On peut bien sûr épiloguer sur les degrés de cette inaccessibilité, qui n'est jamais que relative. Il est toujours possible, en mettant en œuvre les moyens nécessaires, de faire expliquer, traduire ou acquérir un document de cet ordre. Mais il faut bien se rappeler que, quand les moyens mis en œuvre pour avoir une information sont plus pénibles, coûteux ou compliqués que le fait de ne pas l'avoir, sa recherche est abandonnée. Une fois additionnés les coûts occasionnés pour la rendre utilisable, le prix de revient de l'information finale ainsi acquise dépassera de très loin, dans la plupart des cas, la valeur que lui accorde le demandeur ;

- l'information est inintéressante et non pertinente ;
- aucune information n'est retrouvée.

Autant l'utilisateur accepte de payer le prix du service dans le premier cas, autant il sent n'en avoir pas pour son argent dans les cas 2 et 3, se sent nettement volé dans le cas 4. Le cas 5, lui, suscite des sentiments assez mêlés. Dans un premier cas, l'utilisateur pense que l'information recherchée existe et se trouve déçu de n'avoir pas de réponse ; pour lui, le service a mal fonctionné ; certes, il ne sait qui incriminer, des bases de données ou des stratégies de recherche, mais il s'estime floué d'avoir à payer quelque chose. Dans un deuxième cas, l'utilisa-

teur pense que l'information n'existe pas; il est heureux de ne pas avoir de réponse (dans le cas d'une invention nouvelle par exemple) et estime que le prix de la recherche est entièrement justifié; il conserve cependant un léger doute sur la réalité de cette inexistence: ne s'agit-il pas plutôt d'un silence résultant d'une interrogation insuffisamment élaborée?

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 premier cas, éliminer si possible le second, ne se situer dans le traitement qu'avec circonspection, éviter totalement le quatrième et être sûr que le cinquième dit bien ce qu'il veut dire.

... et limiter les coûts

Maintenant que sont posées les données en terme de prix d'achat acceptable, voyons quels sont les moyens qui peuvent limiter les prix de revient.

Le moyen le moins onéreux est de conduire une recherche manuelle avec les moyens du bord. Cette démarche peut parfaitement se justifier: les outils bibliographiques papier permettent une consultation qui peut satisfaire toute une gamme de besoin. Elle n'est cependant possible que lorsqu'il y a parfaite adéquation entre le type de question et l'outil bibliographique disponible, ou lorsque l'utilisateur n'est pas très exigeant et se contente de quelques références. Cette approche n'est valable que si la recherche peut être faite dans un temps relativement court, avec des modalités relativement simples. Dès qu'il faut se déplacer, requérir des autorisations, passer par des procédures administratives compliquées ou avoir recours à des fichiers dont la consultation n'est absolument pas conçue pour répondre à la question posée, le temps passé est tel que, si l'on veut vraiment effectuer la recherche, le recours à l'interrogation des bases de données en ligne est inévitable.

Dans les années soixante, l'expression « explosion documentaire » est devenue la phase de rigueur pour dépeindre à la fois la prolifération des publications et la fragmentation de leurs sujets, devenus de plus en plus spécialisés. A quelle expression faudrait-il recourir aujourd'hui? La situation, en effet, n'a pas cessé

Les critères de sélection s'amenuisent devant un fort afflux de publications trop semblables

d'empirer et retrouver une aiguille dans une meule de foin semble désormais une performance familière presque facile: après tout, l'aiguille est d'une nature différente du foin et peut être détectée avec l'équipement convenable. Actuellement, l'utilisateur se voit plutôt contraint de trouver non plus une aiguille, mais un fil dans une meule de foin, c'est-à-dire un objet presque de même nature. Prenons par exemple le domaine de la recherche industrielle: la compétition y est si forte qu'on peut trouver, dans une recherche brevets portant sur les brevets publiés aux États-Unis de 1984 à 1987, 577 brevets publiés par une même société japonaise, dans la même classe de brevets, pour protéger son domaine d'excellence¹; et quand on sait que ne sont déposés à l'étranger qu'une faible part des brevets japonais... Les critères de sélection s'amenuisent devant un tel afflux de publications trop semblables, parmi lesquelles il faut cependant savoir établir un tri, si l'on ne veut pas être submergé.

Face à cette multiplication de documents, quels sont les moyens dont dispose l'utilisateur pour effectuer une sélection aussi bonne que possible? 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 œuvre d'une batterie de moyens.

Les trois éléments d'une approche stratégique

Une recherche peut être considérablement optimisée par l'adoption d'une stratégie, visant à analyser les différents éléments qui participent à l'interrogation en li-

gne, de façon à les utiliser de la manière la plus pertinente possible.

Circonscrire le champ d'interrogation

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 1970, l'utilisateur des banques et bases de données (BBD) en ligne disposait de quelques dizaines de bases de données, au début des années 80 de plusieurs centaines, et il faut maintenant compter en milliers. De nouvelles couches de connaissances se superposent donc sans cesse à celles déjà nécessaires et qu'il faut continuellement remettre à jour.

Un outil stratégique peut se révéler, quand il existe, fort efficace: il s'agit du « fichier des fichiers² ». Le fichier des fichiers regroupe le vocabulaire présent dans l'ensemble des bases de données présentes sur un même centre serveur; il peut donc permettre la sélection d'un fichier auquel on n'aurait pas pensé *a priori*. En effet, l'utilisateur, même confirmé, interroge de façon constante un nombre limité de fichiers; il en connaît bien la structure et le vocabulaire, et les résultats qu'il obtient sont généralement bons. Devant la multiplicité des bases de données mises à la disposition de chacun, il peut facilement arriver que l'on oublie d'interroger le fichier qui, dans les circonstances présentes, serait le meilleur. Le fichier des fichiers permet de sélectionner le ou les fichiers, où les termes d'interrogation sont les plus employés et où il sera donc possible de trouver le plus grand nombre de réponses satisfaisantes.

1. Il s'agit de la société Fuji photo film, pour la classe de brevets « Compositions photosensibles et leurs supports ».

2. Dans le domaine de la chimie, on peut citer BDI chez SCD, DIALINDEX chez DIALOG, CROSS chez BRS, QUESTINDEX chez ESA-IRS.

R

Le fichier des fichiers fonctionne de la manière suivante : pour chacun des mots du vocabulaire, il affiche le nombre de bases de données dans lesquelles il apparaît, et pour chaque base, le nombre de fois. Les bases de données étant classées par ordre de fréquence décroissante, il est facile de sélectionner le ou les fichiers où le vocabulaire est le plus employé, et donc où il sera possible de trouver le plus grand nombre de réponses satisfaisantes. Prenons un exemple : dans quel fichier trouver les meilleures références sur le recyclage des pétrodollars ? Est-ce dans un fichier économique ou dans un fichier pétrolier ? Une recherche dans un fichier des fichiers donnera le nombre d'occurrences du mot « *pétrodollars* » dans chacun des fichiers dans lesquels il apparaît. Dialindex (DIALOG) permet même une recherche combinatoire et donnera le nombre d'occurrences, par fichier, de l'intersection des deux termes : recyclage et pétrodollars.

Cet outil ne représente cependant 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. Les meilleurs résultats sont obtenus dans le cas d'un vocabulaire non polysémique et associé à des sciences ou techniques fortement interdisciplinaires, telles que l'environnement, la pollution, la sécurité, l'informatique; véritablement situées à la croisée des chemins, ces disciplines seront prises en compte dans un nombre varié de fichiers, pour la plupart déjà connus de l'utilisateur pour d'autres recherches, mais qu'il ne lui serait pas venu à l'idée d'utiliser.

Maîtriser le logiciel d'interrogation

Les procédures d'interrogation s'oublient très rapidement pour qui reste quelque temps sans pratiquer et les quelques secondes de réflexion passées à retrouver la commande exacte finissent, si elles se multiplient, par peser lourd sur le prix de revient d'une recherche. Seule une connaissance approfondie permet d'utiliser au mieux les commandes spé-

cifiques du logiciel. Les systèmes qui offrent des menus (comme Techdata sur BRS) seront bien sûr d'une aide appréciable pour celui qui n'est pas très sûr de sa mémoire et de ses connaissances, mais ils ajoutent à la recherche un surcoût qui peut être très important, allant jusqu'à en multiplier le temps par quatre.

L'utilisateur multifichiers passe tous les temps morts sur un fichier facturé à un prix très bas

Parmi les commandes, il faut insister sur celles qui utilisent au maximum les modalités de tarification du système interrogé : chaque serveur, de par la structure de sa facturation, offre de multiples possibilités de réduire les coûts, notamment par l'utilisation des *parking files* ou fichiers d'attente. A l'Institut français du pétrole (IFP), où le serveur ORBIT est très utilisé — il regroupe les principales bases de données pétrolières —, les *parking files* représentaient, en 1986, 8% du temps de recherche. Les fichiers *parking* sont les fichiers les moins coûteux du serveur; on s'y installe pour tous les temps de transition, comme par exemple pour rédiger une question réclamant une stratégie compliquée et qu'on bascule ensuite dans un fichier beaucoup plus cher. Cela signifie que l'utilisateur multifichiers passe tous les temps morts, inévitables dans une recherche complexe, sur un fichier facturé à un prix très bas, au lieu de rester sur un fichier dont le coût horaire est élevé. On peut aussi s'efforcer d'utiliser systématiquement la base de données du serveur permettant la recherche la moins coûteuse.

Elaborer une recherche efficace

Ce troisième élément consistera par exemple pour l'interrogateur à choisir de manière pertinente les « filtres » qui lui permettront, soit d'élargir, soit de réduire le nombre de références obtenues. Ce pourra être les mots-clés sélectionnés pour l'interrogation et qui, selon le nombre de leurs occurrences, seront élargis ou limités :

- auteurs ou collectivités-auteurs;
- titre ou type de publication;
- langue;
- date;
- classification, codes;
- pondération, quand elle existe...

La pondération apparaît dans les systèmes qui utilisent un grand nombre de descripteurs pour chaque document analysé. Fixons aux alentours de vingt le nombre de descripteurs, à partir duquel le risque de problèmes dus à des phénomènes combinatoires de descripteurs n'ayant entre eux aucun rapport est fortement accru. Les notions alors indexées sont loin d'avoir toutes la même importance : les unes ont une importance majeure, d'autres sont tout à fait mineures, mais seront cependant importantes à retrouver dans certains cas (brevets, application nouvelle, etc.). Il sera donc important de connaître le poids relatif de chaque notion pour une indexation donnée. On donnera pour ce faire des indices différents aux descripteurs principaux, aux descripteurs d'importance moyenne et aux descripteurs accessoires.

L'utilisation de ces « poids » est particulièrement précieuse pour les descripteurs ayant une très grosse fréquence ou lorsqu'on cherche à n'obtenir que des documents d'une très grande pertinence. Elle permet également une approche très intéressante dans l'emploi d'un groupement SAUF. Prenons un exemple : une recherche sur les additifs antidétonnants autres que les additifs au plomb (reconnus comme générateurs de pollution) devrait naturellement s'écrire : additif antidétonnant sauf plomb. Mais le plomb tétraéthyle et les additifs au plomb sont des antidétonnants tellement classiques, qu'il est peu vraisemblable de penser qu'un auteur, décrivant un produit nouveau, ne le comparera pas aux produits anciens; ce type de recherche risque donc d'aboutir à un silence total. En revanche, une stratégie utilisant une pondération retirera le plomb en sujet principal, ce qui laissera subsister les documents où les additifs au plomb ont un rôle mineur, comme celui de faire l'objet d'une comparaison.

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

Une approche stratégique bien menée ne peut être l'apanage que d'un utilisateur confirmé

les commandes les plus classiques et justifier plus que largement 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 : des outils précieux

La démarche est, cette fois, d'une autre nature, puisqu'il s'agit d'étudier le contenu sémantique de la question. Avec l'approche stratégique les éléments étaient explicites, 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. En réalité, seul le spécialiste du sujet possède les connaissances requises pour développer le thésaurus instantané qui lui permettra de mener à bien sa recherche; encore faudra-t-il que ce thésaurus corresponde au 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. Certaines sont conçues directement pour cet usage.

Les thésaurus

Les thésaurus développés par les producteurs de bases de données peuvent pallier un certain manque de connaissances chez l'utilisateur. Matériellement, un thésau-

rus présente une série de termes appartenant à un domaine précis et couvrant chacun un concept ou un ensemble de concepts déterminés; cette détermination s'opérant par le truchement des relations de ce terme avec d'autres termes, relations qui permettent d'en restituer la couverture sémantique.

La première relation allant du spécifique au générique, sans laquelle un thésaurus ne peut exister est une *relation « hiérarchique »* qui remplacera chaque terme dans l'ensemble dont il fait partie. Le problème du choix de la nature hiérarchique des ensembles est le premier problème qu'il faut régler lors de la constitution d'un thésaurus; c'est le problème majeur dont dépendent la nature même du thésaurus et la philosophie de son utilisation. Les relations hiérarchiques permettent de lutter contre le silence découlant d'une question posée à un niveau de spécificité différent de celui des termes du document qui y répondrait: une question posée sur les alliages doit pouvoir

Un thésaurus bien fait permet à un néophyte dans un domaine de formuler correctement une question dans ce même domaine

recupérer aussi bien les articles indexés avec « acier », « bronze » ou « laiton ». Une fois établie, cette relation peut être générée automatiquement par l'ordinateur.

Après cette relation, dite verticale, une deuxième relation sera également nécessaire; horizontale, elle rapprochera des concepts qui sont liés non par rapport hiérarchique, mais par une inférence plus ou moins lointaine, qu'il convient de rappeler à la fois pour l'indexation et pour la recherche de l'information; cette relation s'appelle *relation « associative »* ou de voisinage. Alors que la relation hiérarchique a pour but d'intégrer dans un ensemble les différents éléments qui lui appartiennent, la relation associative établit des ponts entre des ensembles de nature différente. C'est ce réseau de relations avec les autres termes du thésaurus qui définit pour un descripteur son champ sémantique exact. Ces relations ont donc un but précis; elles ne doivent pas être choisies au hasard, ni selon des associations d'idées purement personnelles, mais selon un plan d'ensemble qui replace tous les éléments les uns par rapport aux autres; le souci principal sera d'augmenter la précision de chaque notion et de supprimer les causes possibles de silence dû à une vision trop étroite d'un sujet et qui en ignore certains aspects fondamentaux.

Une troisième relation sera enfin nécessaire: la *relation d'équivalence*. Le problème se pose sur deux plans: problème graphique et problème sémantique. Sur le plan graphique, plusieurs orthographes sont possibles pour un terme et l'on ne retiendra qu'une forme graphique parmi toutes les possibilités. Du point de vue de la sémantique, le problème est d'un ordre différent, puisqu'il s'agit de regrouper sous un même et unique descripteur les différents synonymes et quasi-synonymes disponibles pour exprimer une même notion.

Enfin des notes explicatives précisent l'emploi de certains descripteurs. Un thésaurus bien fait

permet à un néophyte dans un domaine, de formuler correctement une question dans ce même domaine.

Les fichiers dictionnaires

Les fichiers dictionnaires existant en chimie sont un outil précieux pour un vocabulaire qui se caractérise par un grand nombre de synonymes et par 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 à 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.

Co-occurrences et fichiers de citations

Aux moyens directs précédemment décrits, on peut ajouter toute une série de moyens indirects.

L'utilisation des *co-occurrences entre termes* peut ainsi apporter une aide non négligeable à la reformulation des questions. Préfigurant la démarche qu'adopterait un système expert (ZOOM sur ESA-IRS, GET sur INFOLINE, MEMSORT sur QUESTEL), certains systèmes permettent d'associer aux mots-clés les plus significatifs, les autres mots-clés qui les accompagnent le plus fréquemment; la même information peut être obtenue pour les noms d'auteurs et de collectivités-auteurs.

Prenons un 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 ». Nous avons ainsi découvert le nombre important de co-occurrences entre « fractale » et « percolation », ce qui nous a permis de repérer des documents intéressant les spécialistes en production des gisements; utilisant dans leur domaine le terme « drainage » pour les phénomènes de même nature, ceux-ci auraient spontanément associé « fractale » et « drainage » et n'auraient retrouvé aucun document.

Les *fichiers de citations* permettent également d'améliorer les recherches au niveau conceptuel. 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. Le fichier de citations est l'un des moyens les plus sûrs de prendre connaissance d'un vocabulaire nouveau. Il nécessite cependant au préalable d'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 porte l'interrogation, on peut inférer que les documents qui le citeront 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. C'est aussi une solution pour faire un pont entre des publications sur un même sujet, dans des langues différentes. 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 concerné est le mieux armé au départ pour obtenir une belle performance. Mais celle-ci peut tout aussi bien résulter de la recherche imaginative d'un utilisateur confirmé des bases de données.

R

Un type de spécialiste est né : l'expert en interrogation des BBD en ligne

L'effet Matthieu ou l'expert des systèmes

Il est 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 BBD les connaît et les utilise à bon escient, l'utilisateur occasionnel ne pouvant, quelle que soit sa connaissance du domaine interrogé, atteindre les mêmes performances. C'est ce qui a donné naissance à un type de spécialiste, qui se profilait déjà à l'horizon 1974, mais dont on n'était pas alors certain du devenir : l'expert en interrogation des BBD en ligne, qu'il soit à l'origine un intermédiaire ou un utilisateur final reconverti.

Une bonne recherche nécessite le plus souvent de recourir à plusieurs bases de données. Seul l'intermédiaire professionnel, qui a interrogé un nombre de fichiers supérieur (et parfois de beaucoup) à la centaine, peut devenir l'expert capable d'orienter sa recherche sur les fichiers les meil-

leurs 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 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 BBD et des logiciels. Il assistera aux grands congrès documentaires qui, par leurs expositions et revues de produits, sont l'endroit idéal pour acquérir facilement et 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 Parole des talents (Evangile selon Saint-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* ». En d'autres termes, 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 finira par perdre toute activité dans ce domaine.

C'est du moins l'évolution qui se dégage à l'IFP, où la recherche en ligne a débuté en 1974. Avec le triple handicap d'une demande augmentant 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 possible pour maintenir le service demandé, à condition toutefois de s'appuyer sur 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.