

LA RECHERCHE DOCUMENTAIRE AUTOMATISÉE A LA BIBLIOTHÈQUE DU CENTRE SCIENTIFIQUE D'ORSAY

BILAN DE TROIS ANNÉES DE FONCTIONNEMENT
(1977 - 1979)

C'est en mai 1976 que fut installé, à la Bibliothèque du Centre scientifique d'Orsay, un terminal d'accès aux bases de données scientifiques, sous l'impulsion de Mlle Guéniot, directeur de la Bibliothèque de l'Université de Paris-Sud, et grâce au concours du Service des Bibliothèques (Division de la coopération et de l'automatisation) aidée du BNIST (Bureau national de l'information scientifique et technique). Le secrétaire général du BNIST, M. Michel, avait en effet proposé, dès le 28 janvier 1975, lors d'une réunion du Conseil de la Bibliothèque de l'Université de Paris-Sud, la participation de cet organisme à une expérience de recherche documentaire en conversationnel.

Une période transitoire s'étendit de mai à octobre 1976¹ pendant laquelle le personnel put se former à cette nouvelle technique de recherche. Par ailleurs, elle permit de la faire connaître dans le milieu universitaire, les utilisateurs bénéficiant de la gratuité pendant ce laps de temps. Mais, à partir du mois de novembre 1976, une contribution financière fut demandée pour toute recherche. Au seuil de la quatrième année d'existence du Service de la recherche automatisée, il a paru utile de faire le point sur l'utilisation et les types d'utilisateurs des bases de données.

QUELS SERVEURS SONT ACCESSIBLES A ORSAY, ET QUELLES BASES DE DONNÉES ?

Le premier serveur interrogé fut celui de l'ESA (European Space Agency) situé à Frascati en Italie, qui, en 1977, donnait accès à environ une dizaine de bases de données dont les « Chemical Abstracts » (accessible depuis 1969), certainement la plus importante pour nous, étant donné que la chimie est un secteur d'études et de recherches particulièrement actif à Orsay ; « Inspec » : base de données en physique (depuis 1971) et Pascal (depuis 1973), couvrant à la fois les domaines de la physique, de la géologie, de la métallurgie, puis un peu

plus tard, ceux de la chimie et de la biologie. Le nombre de ces bases de données n'a fait qu'augmenter au cours de ces trois années et n'est pas loin d'atteindre la vingtaine à l'heure actuelle. Au milieu de l'année 1977, un contrat fut signé avec les sociétés *Lockheed* et *SDC* (System Development Corporation) et nous permit d'atteindre les gros serveurs américains dont le nombre de bases était bien plus élevé et la couverture beaucoup plus large (par exemple les sciences sociales et économiques, et surtout la biologie qui n'était pas encore accessible sur le serveur européen). Evidemment le logiciel d'interrogation de chacun de ces serveurs était différent : passer de celui de l'ESA (Recon) à celui de Lockheed (Dialog) fut assez aisé, mais une adaptation un peu plus longue s'avéra nécessaire pour celui de SDC (Orbit), complètement différent. Donc, dès la fin de l'année 1977, nous avons la possibilité d'interroger les bases disponibles dans les domaines de la biologie, biochimie, chimie, géologie, métallurgie, physique, et de mettre ainsi à la disposition de nos utilisateurs, quelques millions de références...

QUELS SONT CES UTILISATEURS ?

Ils peuvent être répartis en trois groupes :

- les universitaires proprement dits : enseignants et chercheurs,
- les chercheurs du CNRS et autres organismes d'État installés sur le Campus universitaire ou dans le voisinage immédiat (Gif-sur-Yvette, Thiais, ...),
- les ingénieurs et chercheurs du secteur privé (laboratoires pharmaceutiques, sociétés aéronautiques, sociétés de produits chimiques, etc.).

Dans le premier groupe, il est intéressant de citer les laboratoires qui ont le plus volontiers et le plus souvent recours à cette nouvelle technique bibliographique :

— laboratoires de chimie structurale organique, chimie organique biologique, biologie physico-chimique, structure des matériaux métalliques, physique des solides, synthèse asymétrique, physique nucléaire, biologie expérimentale, physicochimie des rayonnements.

1. Guéniot (Yvonne), Gazier (Denise), Rémy (Danièle).
— Expérience de recherche documentaire automatisée...
In : *Bulletin de la DICA*, 1976, vol. 1, n° 7, p. 3-41.

Dans le second groupe, notons la fréquentation importante du groupe des laboratoires du CNRS de Gif-sur-Yvette, en particulier de l'Institut de chimie des substances naturelles, du Phytotron, celle des chercheurs de l'Institut d'optique situé sur le campus universitaire, des laboratoires de l'École supérieure d'électricité, et d'autres dont il serait fastidieux de dresser la liste.

Le troisième groupe est représenté par des laboratoires comme ceux de l'Institut national des combustibles, Pfizer (laboratoires pharmaceutiques), Quantel (lasers et électro-optique), Finorga (produits chimiques) auxquels nous associerons des entreprises ou établissements publics comme l'EDF à Marcoussis et le CEA.

Si on veut traduire en pourcentage ces diverses catégories d'utilisateurs, on peut dire que :

- 45 % d'entre eux sont des universitaires,
- 35 % des chercheurs du CNRS,
- 20 % des ingénieurs ou chercheurs du secteur privé.

On a toutefois noté une légère augmentation du second groupe en 1979. Quant à la proportion d'utilisateurs du secteur privé, elle reste à peu près constante pour la raison que nous verrons plus loin.

Il faut reconnaître qu'une bibliographie rétrospective ne se fait pas plusieurs fois par an. Il nous arrive de ne revoir les mêmes utilisateurs que deux ans après leur première visite, soit pour une mise à jour de leur documentation, soit parce qu'ils abordent leur problème sous un tout autre angle et doivent donc faire une mise au point, soit aussi parce qu'ils changent complètement de sujet. Ces remarques sont valables pour le premier et le second groupes ; mais pour le troisième, les choses sont un peu différentes. Ils font généralement deux ou trois essais pour tester le système et, comme ils en sont ordinairement satisfaits, ils s'équipent à leur tour d'un terminal et nous perdons ainsi une clientèle intéressante.

Le comportement des utilisateurs est le même dans les trois groupes ; ils reconnaissent tous qu'être présents au moment de l'interrogation en ligne présente un avantage certain, car si le dialogue s'établit entre l'opérateur au terminal et l'ordinateur, il est indispensable qu'il s'instaure aussi entre l'utilisateur final (le chercheur ou l'ingénieur) et l'opérateur (bibliothécaire ou documentaliste). Bien que la question ait déjà été discutée au préalable, il arrive fréquemment qu'une stratégie établie à l'avance soit modifiée au cours de la recherche à la suite d'une intervention du chercheur : par exemple le nombre de références est trop important, il faut affiner le sujet, introduire des descripteurs supplémentaires, ou bien, au contraire, le nombre de références paraît insuffisant, il faut élargir le sujet, ou bien encore, l'indexation de certaines références donne au chercheur l'idée de s'exprimer d'une autre façon... Le dialogue est constant. Un autre aspect du comportement du chercheur est son attitude devant le terminal lui-même ; très peu d'entre eux sont attirés par la manipulation elle-même, ils préfèrent « regarder » les opérations se dérouler sous leurs yeux étonnés ou quelquefois sceptiques.

LA FRÉQUENCE D'UTILISATION DES BASES DE DONNÉES

Le Service de la recherche automatisée n'interrogeant que des bases de données de sciences pures (les bases de sciences médicales, pharmaceutiques, sociales ou économiques étant exploitées dans les bibliothèques des autres centres de l'Université de Paris-Sud), le nombre de recherches et de temps d'utilisation (fig. 1 et 2) paraîtra sans doute assez modeste, encore qu'on constate une progression d'année en année.

Sans doute faut-il reconnaître qu'il existe une certaine méfiance vis-à-vis de cette nouvelle technique. Méfiance à propos de la fiabilité des systèmes : est-on sûr d'avoir vraiment tout ce qui est paru sur le sujet qui vous intéresse ? Méfiance sur l'exploitation éventuelle qui pourrait être faite de l'interrogation des bases de données : le sujet de la recherche ne risque-t-il pas d'être connu ? Le secret est-il bien gardé ? Il est plus facile de rassurer l'utilisateur sur ce dernier point que sur le premier, car l'ordinateur étant vraiment anonyme, les chercheurs se rendent bien compte qu'espionner un tel système dépasse la commune mesure. Par contre, les convaincre que rien ne leur a échappé est déjà plus malaisé.

Car il est bien connu que toute indexation est subjective. La personne qui a analysé le document, si compétente soit-elle dans le sujet traité, a introduit une part de subjectivité dans son analyse. L'opérateur et le chercheur, à leur tour, vont choisir subjectivement les mots-clefs qui leur paraissent les plus appropriés pour la recherche envisagée mais qui ne seront pas nécessairement ceux retenus par l'analyste. Entre le document et le chercheur, il n'y a pas seulement un intermédiaire, comme c'est le cas dans la consultation d'un catalogue ou d'une bibliographie mais deux.

C'est pourquoi la stratégie de recherche doit-elle être soigneusement établie avant l'interrogation en ligne et, au temps de la recherche proprement dite, doit être ajouté le temps de préparation de la question qui peut varier de quinze minutes à trois heures, en particulier dans les domaines de la chimie et de la biologie. Le gain de temps peut être appréciable lorsque le chercheur possède déjà deux ou trois articles sur le sujet. L'indexation peut ainsi en être connue et donner un excellent point de départ à la recherche elle-même.

Les courbes représentées sur les figures 1 et 2 permettent, d'une part, de faire une étude de l'évolution du nombre de recherches et du temps passé à conduire ces dernières au cours d'une année, et, d'autre part, de faire une comparaison sur les trois années. Si nous analysons leur tracé mois par mois, nous constatons que les périodes d'activité et de ralentissement suivent très exactement le rythme de la vie universitaire, ce qui ne saurait surprendre puisque, nous l'avons vu, les universitaires et les chercheurs du CNRS représentent la majorité des utilisateurs.

C'est ainsi que les courbes s'effondrent en avril, du fait des vacances pascales, et en juillet-août, du fait des vacances universitaires. On enregistre également un faible taux d'interrogation pendant les

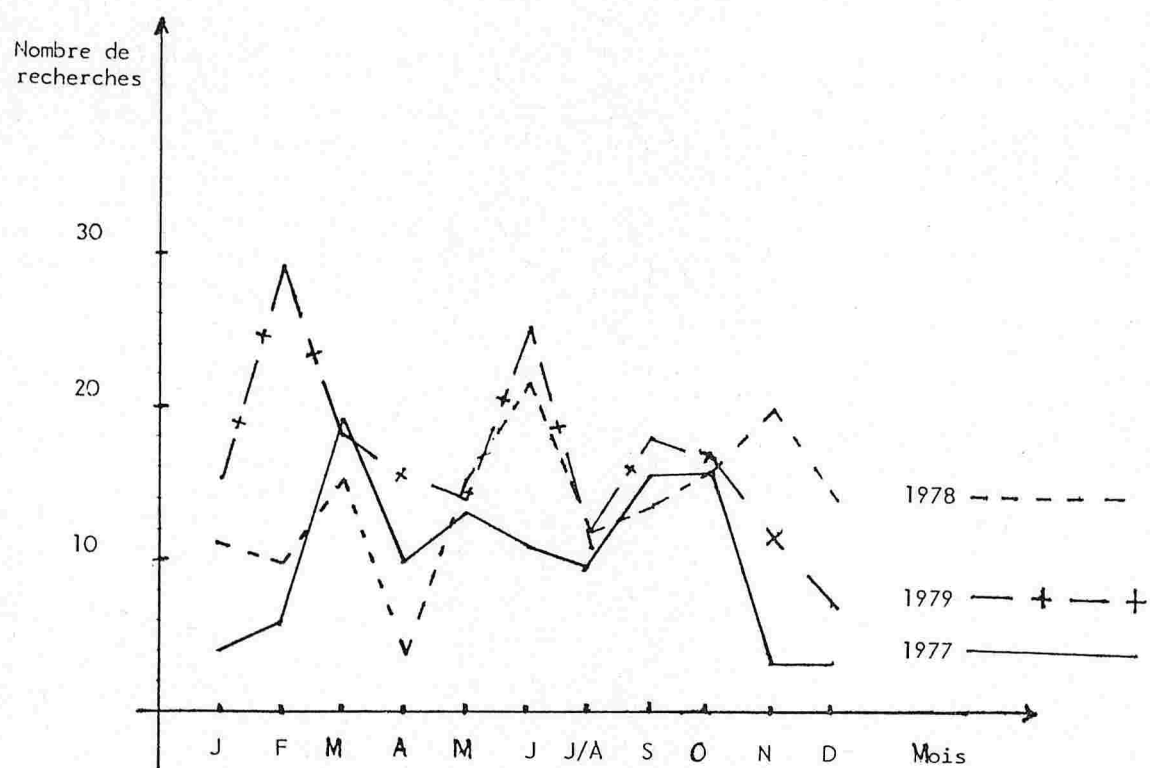


Fig. 1

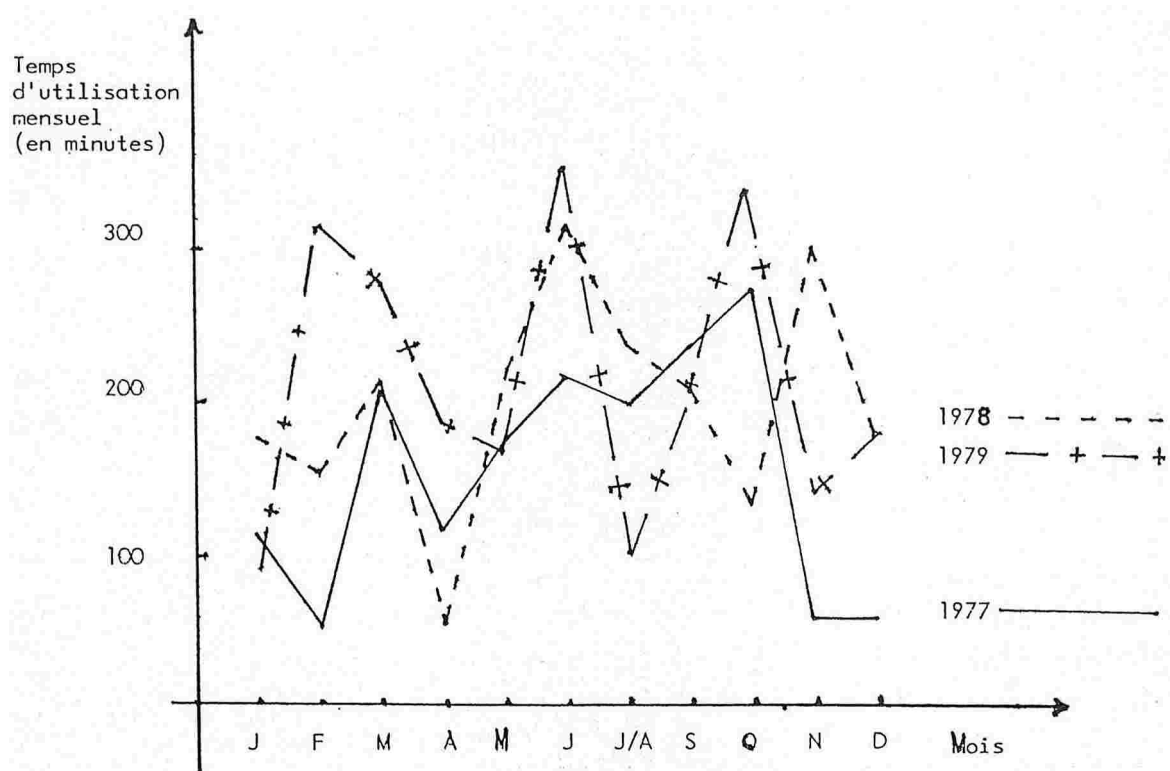


Fig. 2

mois de janvier et décembre et ceci pour des raisons essentiellement budgétaires : en janvier, les sommes attribuées à chaque laboratoire ne sont pas fixées avec certitude tandis qu'en décembre le budget est épuisé d'où un nombre peu élevé de recherches.

Les temps forts, en revanche, se situent en février-mars, en mai-juin, puis en septembre, octobre et novembre. Au cours de la première période, nous assistons même à un phénomène de pointe annuelle : en février pour 1979 alors qu'il est en mars pour 1977 et 1978. Pourquoi cette soudaine montée en début d'année ? D'une part, les laboratoires universitaires et ceux du CNRS sont à même de répartir leur budget. De nouveaux programmes de recherche sont lancés : il va donc être intéressant de savoir qui travaille sur le même sujet et d'aborder éventuellement la recherche sous un angle différent de ceux adoptés jusqu'alors. D'autre part, c'est la pleine période de soutenance de thèses d'où la nécessité absolue de faire une sérieuse mise au point de sa bibliographie afin d'avoir une bonne vue d'ensemble de la littérature parue tout récemment.

En mai-juin, on note également une nette remontée qui culmine en juin. La fin de la période d'enseignement approche et, il est possible de consacrer plus de temps à la bibliographie. Il faut déjà penser aux nouveaux thèmes qui pourront être proposés à la prochaine rentrée universitaire. De plus, la période d'été est riche en congrès nationaux et internationaux, donc favorable à des consultations de fichiers qui permettent ainsi de mieux connaître l'orientation des recherches des participants et de vérifier l'originalité de ses propres communications. Ceci se traduit par le style des questions posées à cette époque qui se révèlent être essentiellement des bibliographies d'auteurs.

La dernière période d'activité s'étale sur trois mois (septembre, octobre, novembre) qui se situent à peu près au même niveau. L'augmentation du nombre des interrogations est due aux causes suivantes : de nouveaux sujets de recherche sont distribués, sujets de DEA, de thèses de 3^e cycle ou d'État. C'est aussi l'époque de la rédaction des rapports annuels au CNRS et les chercheurs veulent connaître rapidement les dernières références parues sur le programme de recherche qui les concerne.

L'analyse des figures 3 et 4 permet de voir l'évolution du nombre des utilisateurs et du temps de recherche au cours de ces trois dernières années.

Sans doute l'augmentation n'est-elle pas vertigineuse mais elle est quand même en constante progression, ce qui est rassurant et tend à prouver que ce nouveau mode de documentation fait son chemin.

A propos du temps passé pour exécuter les recherches, on peut faire l'observation suivante : en 1977, le temps moyen d'une recherche était de 16 mn, en 1978 de 14 mn et en 1979 de 13 mn ; évidemment la pratique des stratégies intervient, mais aussi l'utilisation plus fréquente de la demande en différé qui permet à un moindre prix d'obtenir un maximum de références sur un thème donné.

LES BASES DE DONNÉES LES PLUS CONSULTÉES

Le tableau dressé ci-dessous tient compte des bases les plus interrogées mais aussi des serveurs consultés.

D'après ce tableau, on peut conclure très facilement que les Chemical Abstracts sont les plus consultés (environ 50 %), ceci pour plusieurs raisons : les domaines couverts par ces abstracts sont très variés. Lorsqu'on regarde la liste des sections dépouillées par les Chemical Abstracts, on s'aperçoit que naturellement la chimie qu'elle soit organique, minérale, physique, y est représentée, mais aussi la métallurgie, la minéralogie, même la biologie et la physiologie (hormones et enzymes). Une autre raison intervient, c'est la difficulté de consultation des index des Chemical Abstracts, et de manipulation d'un grand nombre de fascicules avant d'obtenir la ou les références cherchées, d'où l'avantage de parcourir un listing qui vous donne, avec un encombrement minimum, le signalement complet des citations qui vous intéressent.

PASCAL (Bulletin signalétique du CNRS) apparaît en seconde position, immédiatement suivi par BIOSIS (Biological Abstracts) ; mais il faut tenir compte du fait que PASCAL n'est apparu qu'en 1977, et que, de plus, étant pluridisciplinaire, il fausse un peu les résultats ; malgré tout, l'augmentation des interrogations sur ce fichier est très nette, car bien souvent elles complètent celles qui sont exécutées sur Chemical Abstracts, sur Inspec ou Biological Abstracts ; ceci est dû en partie à l'origine des documents dépouillés par le CNRS, notamment ceux provenant des journaux européens. Et n'oublions pas que le CNRS fournit

Nombre de recherches sur trois ans

Bases de données Serveurs							
	BIOSIS	CA	INSPEC	METADDEX	NASA ²	PASCAL ²	
ESA/IRS ¹	29	96	42	7	13	68	255
LIS ¹	35	92	6	3			136
SDC	3	13					16
	67	201	48	10	13	68	

1. IRS : Information Retrieval Service
LIS : Lockheed Information Systems

2. Les fichiers encadrés ne sont interrogeables que sur le serveur européen ESA/IRS.

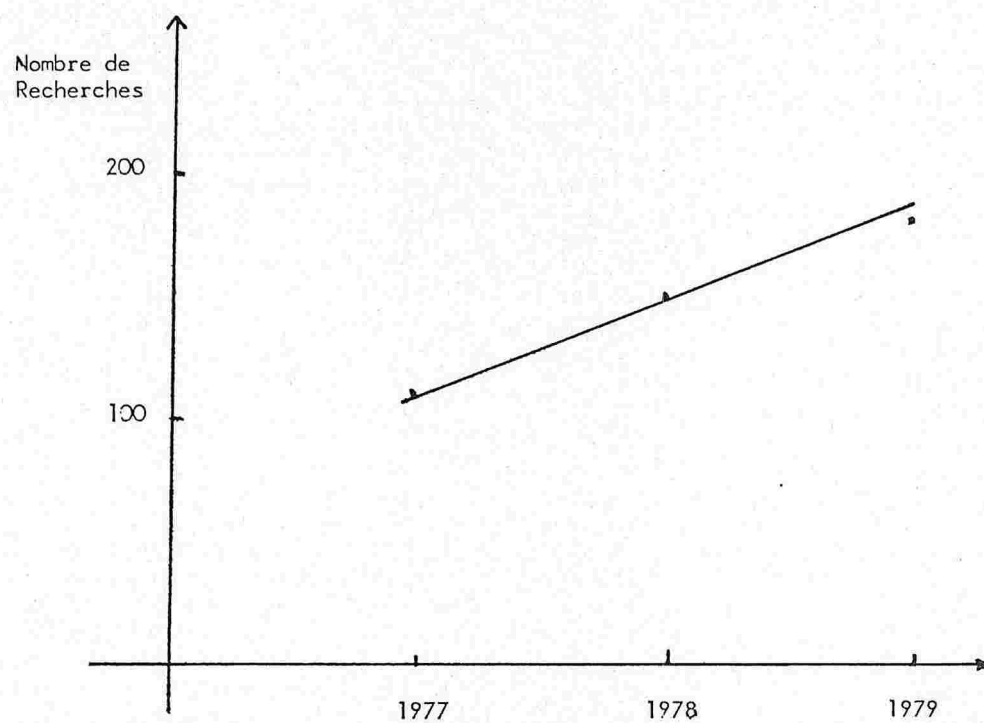


Fig. 3

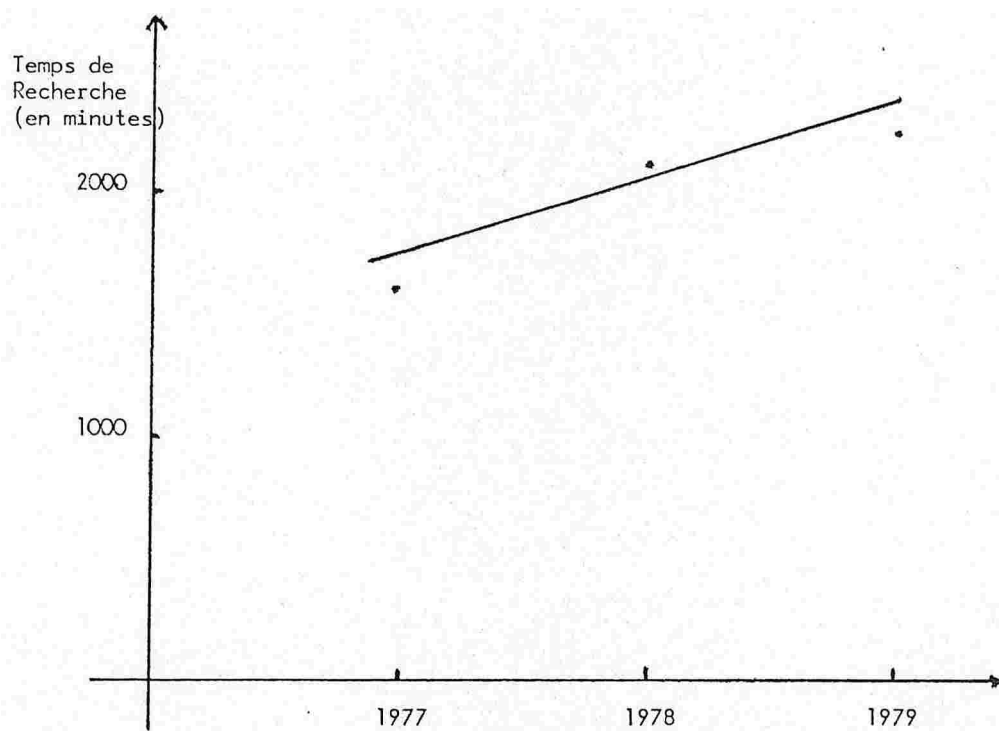


Fig. 4

les documents primaires, condition primordiale pour les utilisateurs, car donner une liste « alléchante » de références bibliographiques n'est pas tout. Encore faut-il être capable de procurer au client les articles eux-mêmes. Si nous ne sommes pas à même de le faire, nous nous heurtons parfois à l'incompréhension de certains d'entre eux, ce qui évidemment n'est pas une bonne publicité pour le développement de cette nouvelle technique. Sans doute, le fonctionnement plus intense du prêt inter-bibliothèques permet dans une certaine mesure de remédier à cette carence, mais les délais, plus ou moins longs, d'obtention des documents sont un autre handicap.

Pourquoi un serveur plutôt qu'un autre ? Comme il a déjà été dit un peu plus haut, les logiciels de l'ESA/IRS et de Lockheed/LIS sont très voisins, d'où un nombre d'interrogations plus important sur ces deux serveurs. Pour BIOSIS, la période de couverture étant plus grande sur LIS, et l'apparition de la base plus tardive sur IRS, le phénomène s'explique très bien. Pour les Chemical Abstracts, il faut signaler une nette augmentation des questions posées sur IRS en 1979, alors que celles réalisées en 1977 et 1978 l'étaient plutôt sur LIS, ceci est dû surtout à l'apparition de nouvelles clés d'accès sur IRS (numéro de registre, formule moléculaire) n'existant alors que sur le serveur LIS. Le système européen a été le plus consulté pour les bases de physique (Inspec), car à possibilités d'accès égales, le choix s'est porté naturellement sur l'ordinateur le plus proche.

Par ailleurs, la présence des fichiers NASA et PASCAL sur IRS seul, justifie la différence du nombre d'interrogations en faveur de celui-ci.

On peut dire que, dans l'ensemble, le bilan est positif, car le nombre des utilisateurs a augmenté lentement mais régulièrement au cours de ces trois années. La fidélité de certains usagers nous conduit même à envisager une nouvelle forme de service. En effet, après avoir obtenu une bibliographie rétrospective, le besoin de la mettre à jour mensuellement se fait sentir : c'est pourquoi, pour répondre aux demandes qui nous sont faites, nous étudions actuellement la possibilité de réaliser automatiquement des profils.

L'apparition d'un serveur national français TÉLÉ-SYSTÈMES, devrait sensibiliser un plus grand nombre d'usagers, en particulier les chimistes et biochimistes. Le logiciel DARC, mis au point par l'équipe du Pr Dubois à Paris VII, doit leur permettre d'accéder aux structures et sous-structures d'environ 400 000 composés chimiques, et certainement beaucoup plus dans un proche avenir. Les références correspondant à ces structures se trouvent dans la partie CBAC (Chemical Biological Activities) des Chemical Abstracts, accessible depuis 1965 par le logiciel MISTRAL sur TÉLÉSYSTÈMES. D'autres bases de données françaises sont chargées sur ce même serveur (BIPA, CANCERNET, EDF, ...) ce qui permettra de toucher un public beaucoup plus vaste, et facilitera la tâche des opérateurs, dans la mesure où il est possible d'utiliser un logiciel français.

Danièle RÉMY

*Responsable du Service
de recherche documentaire automatisée*