

MENTHE ET KETCHUP. Quel que soit l'ingrédient dominant, la sauce européenne a un goût bizarre. Il est vrai qu'on est dans le domaine très particulier de l'information chimique ; un monde où les Européens peuvent afficher des réalisations triomphalistes et des prestations pointues, et, tout en même temps, s'accrochent à des positions défensives face à l'omnipotence américaine. Le salut viendra-t-il de l'Europe ? On peut en douter au vu du bilan, très nuancé, dressé par Monty Hyams. Il est vrai qu'il s'agit là d'un point de vue très très « british »...

Le secteur de l'information chimique constitue un marché à caractère vraiment international ; les Européens y jouent un rôle très important, surtout dans les créneaux dont j'ai choisi de vous parler aujourd'hui. Ces dernières années ont vu s'accroître de façon fantastique l'usage des systèmes d'information en chimie utilisant un code topologique — comme CAS-ONLINE, DARC Télé systèmes et l'ensemble des programmes de Molecular design.

Codages topologiques

Les systèmes codent les structures chimiques en déterminant la configuration de la molécule et les interconnexions entre les atomes.

* La version anglaise de cette communication a été publiée dans *Aslib proceedings*, vol. 39, n° 5, mai 1987. La traduction a été effectuée avec la collaboration de Bernard MARX.

Ils vous permettent de poser les questions et de recevoir les réponses sous la forme de diagrammes de structure chimique, qui sont, bien sûr, le langage non-ambigu normalement utilisé par les chimistes. Cependant, cette approche ne permet pas actuellement de résoudre le problème du codage des structures dites « formule de Markush » (formule générique que l'on trouve dans les brevets de chimie).

Les systèmes de codes topologiques comme DARC et CAS-ONLINE sont conçus pour coder individuellement des composés dans la banque de données. On peut effectuer des recherches sous-structurales en utilisant des notions génériques de recherches, mais la banque de données dans laquelle la recherche est effectuée ne comporte que des composés individuels ; dans le cas de CAS-ONLINE, on cherche l'ensemble du CAS registry qui contient plus de 6 millions de composés. Cependant, une seule revendication dans un seul brevet de chimie peut facilement couvrir 6 millions de composés et il existe de nombreuses formules de Markush qui peuvent couvrir une infinité de composés.

Prenons le cas du substitut R (schéma 1) ; celui-ci, entre autres possibilités, pourra être un aralkyl. La longueur de la chaîne alkylique n'est pas précisée, ni le type de substitution, ni le nombre de substituants présents. Un système tel que CAS-ONLINE devrait théoriquement prendre en compte chacun des composés possibles à partir de cette formule de Markush et les coder, ce qui est impossible.

CAS tourne cette difficulté en codant seulement les composés qui ont été précisément découverts et

décrits dans les caractéristiques de brevets et, plus récemment, dans les revendications spécifiques. De tels raccourcis ne sont pas utiles à la communauté scientifique utilisatrice des brevets ; recherchant les innovations et les contrefaçons, elle a besoin d'avoir la vision la plus large possible.

Au cours des vingt-cinq dernières années, les structures de Markush ont, dans la littérature des brevets, été codées de façon satisfaisante par seulement deux organismes, IDC à Francfort avec le code Gre-mas, et Derwent à Londres qui utilise aussi un code de fragmentation. Il n'existe pas d'autres fichiers satisfaisants sur le marché international. Dans les deux cas, le codage tenait le bon bout pour prendre en compte toutes les possibilités, mais, du fait du sous-codage et de l'incapacité à définir exactement les points d'établissement des substituants, le nombre d'erreurs est souvent vraiment excessif et rend fastidieux l'affichage des résultats. D'autre part, le codage est très complexe, demandant une très grande maîtrise de la part du chercheur. Derwent a récemment apporté des améliorations, assurant ainsi sa prédominance sur le marché.

En travaillant en liaison avec Télésystèmes à Paris, Derwent a maintenant amélioré une extension du système DARC, lequel peut aussi être utilisé pour les formules de Markush et les fichiers sont actuellement rechargés pour des tests. Il faut environ une heure pour analyser chacun des brevets contenant une structure de Markush, et il y en a environ 400 par semaine, à travers le monde.

Le schéma n° 2 montre la première des deux pages de l'analyse de structure présentée dans le schéma n° 1, de manière à donner une idée du travail effectué en amont. On a appelé G 7 une des variables de R et on peut voir combien il y a de possibilités. Au passage, je signale que des termes tels que CHK (alkyl) sont connus sous le nom de super-atomes, un concept fondamental du système DARC-Markush. Un crayon optique et un micro-ordinateur IBM-XT suffisent pour transférer les données sur l'ordinateur central de Télésystèmes, à Valbonne dans le sud de la France.

N'importe quel brevet générera un écran des structures de base, en l'occurrence celle présentée par le schéma n° 3, montrant les points de branchement des différents groupes. Il générera aussi, pour

« Les brevets de chimie s'insèrent dans un système de revendications complexe à interpréter, même pour les spécialistes les plus compétents en la matière, et qui constitue un obstacle légal »

chacune des variables, un certain nombre d'écrans auxiliaires, celui présenté par le schéma n° 4 montrant la variable G 7 du substitut R dans la structure de Markush. Pour effectuer une recherche, il suffit simplement d'entrer sur un écran de micro-ordinateur la représentation graphique de la structure générique recherchée. Compte tenu du codage du fichier à plusieurs niveaux, il est possible de retrouver un brevet en effectuant une recherche large lorsque l'on souhaite s'informer sur les nouveautés, ou en faisant une recherche plus pointue, lorsque l'on cherche à repérer une contrefaçon.

IDC, un consortium d'entreprises chimiques, surtout allemandes, qui utilise le code Greimas, et qui est le seul autre organisme à coder les structures de Markush, a décidé de continuer à l'utiliser, uniquement pour permettre à ses codeurs et à ses chercheurs de générer plus facilement ces codes complexes, en utilisant un système d'entrées graphiques. Ils prévoient d'utiliser le système Gensal, mis au point par Lynch, professeur

à l'Université de Sheffield, pour développer ces codes à travers des tables de connectivité.

Pourquoi a-t-il fallu mettre en place un système de recherche aussi compliqué ? Parce que les brevets de chimie s'insèrent dans un système de revendications complexe à interpréter, même pour les spécialistes les plus compétents en la matière, et qui constitue un obstacle légal.

La difficulté tient à ce que, à la différence des brevets américains, qui ne sont publiés qu'après un examen approfondi, les demandes déposées en Europe et au Japon sont publiées telles quelles et à l'état brut. Pire, il est nécessaire d'attendre plusieurs années après le dépôt d'un brevet avant de savoir ce qui peut en être revendiqué, afin de décider si l'on s'expose à contrefaçon. Une autre difficulté tient à ce que le contenu des spécifications d'un brevet européen n'est pas obligatoirement rédigé en anglais ; par ailleurs, 35 % des brevets de chimie de base sont en japonais.

Derwent, à Londres, et IDC, à Francfort, sont les seuls organismes à travers le monde susceptibles d'intervenir sur cette composante fondamentale du marché international de l'information chimique.

Analyse statistique

Les services d'information fournissent en général des résumés d'articles, de rapports et de brevets, tous documents censés correspondre aux besoins des services opérationnels. Or, ce que ces derniers veulent, en fait, ce sont des analyses comportant une évaluation, faisant le point sur l'évolution de la recherche, de la technologie, de la concurrence, de la stratégie de produits. Il s'agit là d'une donnée fondamentale dans le secteur extrêmement compétitif et axé sur la recherche qu'est l'industrie chimique. L'analyse statistique, utilisée en liaison avec la recherche par sous-structure, pourrait ainsi permettre d'identifier les types de composés auxquels on s'intéresse et les lieux de cette recherche ; elle permettrait aussi de dresser la carte des nouvelles relations activité-structure qui sont en train d'apparaître.

De nombreux organismes publient régulièrement des rapports donnant les résultats de leurs analyses et de leurs prévisions, tels feu l'OTAF (Office of technology assessment and forecast) qui publiait les *Reports of the US patent and trademark office*. Derwent s'est aperçu que, ce qui intéresse les abonnés, c'est de pouvoir récupérer pour leur usage propre les analyses qui les concernent et il a enregistré une formidable demande en ce sens sur ses fichiers de brevets en ligne. Derwent a néanmoins considéré que l'installation, en complément d'une banque de données, d'un service de traitements statistiques totalement opérationnel était pour l'instant impossible pour un ensemble de raisons diverses, tant économiques que techniques.

La plupart des programmes de traitement statistique sur gros ordinateurs, tel SAS, tournent très vite, à condition d'utiliser autant de mémoire d'unité centrale qu'il est besoin pour une analyse. Cette caractéristique est à l'opposé des besoins d'un système d'information opérationnel en ligne. Si beaucoup d'utilisateurs faisaient faire en même temps leurs analyses, tous les autres seraient blo-

« La possibilité de réaliser des analyses statistiques ouvre désormais un nouveau marché à l'information chimique, celle des services cousus d'or des ventes et de la prévision et non plus celle des services jumeaux de la recherche et du développement, souffrant d'une pénurie chronique en matière de finances »

qués jusqu'à la fin du traitement. Par ailleurs, alors qu'il est possible, avec le temps, de planifier les coûts informatiques de la recherche en ligne, en appliquant un tarif horaire de connexion, les coûts d'une analyse statistique sont déterminés par le type de traitement demandé. La solution consiste à limiter le volume des analyses autorisées. La commande GET de Pergamon-Infoline en est un exemple, mais la taille limite de 200 lignes pour chaque analyse en compromet quelque peu l'efficacité.

Télé systèmes-Questel propose un service plus restreint, MEMSORT, sur son nouveau logiciel Questel Plus, mais son efficacité décroît en fonction du nombre de données à traiter. Alors même que les meilleures performances n'interviennent que sur des corpus réduits (facteur occurrence x données : 500 x 1 000), Télé systèmes a affiché tout l'intérêt qu'on pouvait retirer de ce service, soulignant l'aide que de tels systèmes pouvaient apporter à la communauté des utilisateurs.

Derwent, quant à lui, a opté pour une approche complètement nouvelle : le télé-déchargement de données à partir d'un serveur en ligne, leur analyse en différé, sur un micro-ordinateur. Le programme, PATSTAT, était à l'origine basé sur DBASE II. Il a maintenant été réécrit en PASCAL et tourne 20 fois plus vite que dans la version de départ. A 1 200 bauds, il est possible, en une heure, de télé-décharger soit 1 000 enregistrements de brevets Derwent contenant chacun 10 paramètres à analyser, soit 10 000 enregistrements avec chacun un paramètre — celui du cessionnaire du brevet. Même avec des recherches larges, couvrant de vastes domaines où le volume de données à traiter est plus important, les micro-ordinateurs peuvent effectuer leurs analyses à l'intérieur d'une capacité standard de stockage de 20 megabytes.

Le schéma n° 5 présente un exemple d'analyse PATSTAT, un histogramme des brevets sur les anticorps monoclonaux, retraçant

l'activité de recherche des entreprises les plus actives au cours de la période définie par l'année de priorité de l'invention. L'information sur l'activité des entreprises en matière de dépôts de brevets est complétée par les données disponibles sur le monde des affaires, en interrogeant PROMT et CIN (Chemical industry notes). Ainsi, le bond en avant réalisé par GREC (Green Cross) en 1983 tient aux accords de coopération en matière de recherche et développement intervenus l'année précédente entre cette entreprise et l'Université de Californie. L'inactivité de Toray (TORA) en 1981 trouve son explication dans les conventions sur les brevets passées avec FUJI, dont la présence se manifeste à partir de l'année de priorité 1982. Du Pont n'apparaît dans le champ qu'en 1983, une année après avoir annoncé l'affectation de crédits sensiblement majorés au secteur de la recherche-développement. Le ralentissement du rythme des brevets déposés par Hybritech's (HYBR) s'explique peut-être par l'annonce d'un certain nombre de produits nouveaux pour 1984-1985, ce qui permet d'augurer un renversement de sa politique, de la recherche vers le développement et la production.

Une retombée négative de l'approche fondée sur l'utilisation du micro-ordinateur tient à ce que, une fois vendu le programme, les seuls revenus possibles pour le producteur de la banque sont ceux qu'il peut retirer du télé-déchargement. Dès lors que PATSTAT, qui est vendu pour à peine 550 \$, peut aussi servir pour analyser d'autres banques de données, Derwent ne pourra même pas, dans l'état actuel des choses, récupérer des profits grâce au télé-déchargement.

En définitive, l'un des avantages les plus importants de l'analyse statistique, aux yeux du spécialiste de l'information, est de le mettre à même de sélectionner uniquement les documents pertinents, ceux qui contiennent au moins un certain nombre de fois un terme de recherche donné, ou d'établir chaque mois une analyse statistique de tous les nouveaux résumés d'articles publiés dans un domaine, triant les termes les plus fréquents et fournissant les moyens de repérer immédiatement les évolutions de la recherche.

Qui plus est, cette possibilité de réaliser des analyses statistiques ouvre désormais un nouveau marché à l'information chimique, celui

des services cousus d'or des ventes et de la prévision et non plus celui des services jumeaux de la recherche et du développement, souffrant d'une pénurie chronique en matière de finances.

Mécanismes réactionnels

Les mécanismes réactionnels forment un élément fondamental, mais quelque peu négligé, de l'information chimique. Jusqu'à une date récente, la recherche par représentation graphique restait du domaine de l'impossible. Au cours des dernières années, on a enregistré une activité effervescente en la matière. Il existe, à l'heure actuelle, au moins quatre logiciels de recherche disponibles dans le commerce ; ceux-ci sont conçus pour le stockage et la recherche en local de réactions chimiques à partir d'entrées et de sorties graphiques. Il s'agit de logiciels suivants : REACCS (Molecular design, Californie), SYNLIB (à l'origine conçu par Smith Kline & Beckman, États-Unis, maintenant aux mains du secteur privé), ORAC (Université de Leeds, Grande-Bretagne), DARC-RMS (Télé systèmes, France). REACCS fut le premier à apparaître sur le marché et possède un avantage marqué par rapport à ses concurrents, car sa banque de données est la plus importante et la plus complète. Par ailleurs, REACCS peut être utilisé sur bon nombre de mini-ordinateurs alors qu'actuellement ORAC ne tourne que sur VAX et DARC-RMS sur IBM.

Par contre, REACCS ne comporte pas certaines des caractéristiques attrayantes possédées par ORAC et DARC-RMS, en particulier la possibilité de dresser la carte, atome par atome, des centres de réaction. C'est pour cette raison que REACCS apparaît actuellement peu adapté au traitement de certains types de réaction, mais cette insuffisance devrait être prochainement réduite.

On a dit qu'ORAC possédait le meilleur logiciel ; celui-ci suscite déjà un vif intérêt en Europe, mais n'a pas jusqu'à présent réussi à pénétrer les marchés américain et japonais. Ses principaux handicaps tiennent à la dimension de sa banque de données (15 000 enregistrements seulement) et à sa réputation de produit conçu et géré par une petite équipe d'universitaires faiblardes en matière de marketing.

SYNLIB a rencontré un certain succès tant aux États-Unis qu'en Eu-

rope. Mais son logiciel apparaît moins sophistiqué que ceux de ses concurrents. Il a lui aussi connu des difficultés pour cause de gestion et n'est actuellement plus aux mains de Smith Kline. Sa banque, qui comprend quelque 30 000 enregistrements, est critiquée pour la place trop importante qu'elle donne à la chimie alicyclique.

DARC-RMS comporte beaucoup de caractéristiques semblables à ORAC, y compris d'excellentes possibilités de recherche par sous-structure. Cependant, ne comportant pas de banque de données, il n'a, même en France, que faiblement pénétré le marché. Les logiciels ont besoin de données pour effectuer une recherche. Les banques de données sur les réactions, véritablement significatives, sont celles, effectives ou projetées, émanant d'ISI, de CAS et de Derwent.

Les *Current chemical reactions* d'ISI, lancées en 1979, comportent, à ce jour, 28 000 entrées. Elles ne couvrent qu'une centaine de titres et excluent la littérature sur les brevets. Elles doivent bientôt faire l'objet d'une sortie sous forme de bande magnétique, réservée cependant à la recherche à domicile et liée exclusivement à l'utilisation de REACCS.

« Les chimistes ont besoin de savoir comment ils pourront plus facilement synthétiser la substance X : en d'autres termes, ils ont besoin d'une formule »

Le service des *Chemical abstracts*, prévu pour 1988-89, ne couvrira non plus, sans faire de rétrospectif, qu'une centaine de titres ; on prévoit cependant qu'il intégrera 300 000 réactions par an, incluant toutes les réactions intermédiaires d'une synthèse isolée méconnaissant les règles de l'innovation. Il s'agit là d'une décision surprenante, et d'un bon exemple de surcharge d'information. On a le sentiment que les tout-puissants *Chemical abstracts*, réalisant tout d'un coup qu'ils avaient été absents sur ce marché, se sont précipités pour annoncer leurs projets, sans avoir fait auparavant l'étude de marché qui est indispensable.

En Grande-Bretagne, Derwent s'est adressé à Fraser Williams pour créer à partir de sa base de données un fichier graphique utilisant des tables de connectivité,

de façon à permettre à tout utilisateur de travailler après conversion avec l'un ou l'autre des logiciels graphiques qu'on vient d'évoquer.

Un des problèmes de fond posés par ces services de réaction est qu'ils n'offrent que des références bibliographiques. Les chimistes ont, la plupart du temps, besoin de savoir comment ils pourront plus facilement synthétiser la substance X ; en d'autres termes, ils ont besoin d'une formule. De ce point de vue, le nouveau service créé par Pergamon, connu à titre provisoire sous le nom de CHEM-QUEST, mais dont l'appellation est contestée par Questel, semble intéressant. Il a été lancé lors du dernier congrès de l'American chemical society à Anaheim, et constituera un des développements de la base des Fine chemicals directory, récemment rachetée à Fraser Williams. D'après Pergamon, ce service permettra d'entrer la représentation graphique d'une molécule ou d'une de ses parties. Le système est censé répondre en indiquant un cheminement et en guidant l'utilisateur pour obtenir les produits de départ. Il prendra également en compte la vitesse de réaction et son efficacité, ainsi que les coûts des produits de départ.

Les autres secteurs où les chimistes ont besoin de réponses simples — banques de données factuelles sur les propriétés, information spectroscopique, etc. — présentent des caractéristiques proches de celles des banques sur les réactions. Il existe dans ce domaine des banques en ligne, mais elles ne sont que faiblement consultées — la seule solution acceptable sur le plan commercial consisterait pour les vendeurs en des banques de données en local, avec un logiciel spécial ; ainsi verrait-on se constituer un nouveau créneau d'exploitation.

L'Europe et l'information en ligne

Au tout début de la recherche en ligne, le docteur G. Anderla de la Commission européenne eut la clairvoyance de réaliser qu'il était

« *EURONET appartient désormais au passé et les différents facteurs qui l'on conduit à l'échec sont encore présents, si bien qu'à l'heure actuelle le marché européen est fragmenté sans espoir d'amélioration et toujours dominé par les Américains* »

indispensable de procéder à un effort de coopération au niveau européen, pour résister à la domination américaine sur le marché. Le projet EURONET fut donc mis en place. Il avait deux objectifs principaux : éviter la duplication de banques de données à l'intérieur du réseau ; abaisser les coûts de télécommunications et les uniformiser en passant des accords entre les différents pays ; l'intégrité de ces services ne devait pas être remise en cause du fait des liaisons transatlantiques.

Ces objectifs n'ont pas été remplis pour cinq ensembles de raisons diverses : les serveurs n'ont pu se mettre d'accord sur la question de savoir qui exploiterait les banques les plus rentables ; on ne disposait d'aucun logiciel pouvant faire l'affaire, mais on ne disposait pas non plus des fonds nécessaires à un développement ; les barrières linguistiques freinaient les efforts de formation et de promotion entre les différents pays, ainsi en est-il allé de COMPENDEX, financé par l'ensemble de la communauté et disponible uniquement en Allemagne. Le nombre de banques n'a jamais atteint la masse critique pour permettre la recherche croisée par fichiers ou pour justifier l'apprentissage d'un nouveau langage de commande. Les communications EURONET n'ont jamais eu la puissance et la fiabilité nécessaires et les réseaux nationaux de télécommunications ont manqué complètement d'enthousiasme pour mettre en œuvre une politique de concessions mutuelles. EURONET appartient désormais au passé et les différents facteurs qui l'ont conduit à l'échec sont encore présents, si bien qu'à l'heure actuelle le marché européen est fragmenté sans espoir d'amélioration et toujours dominé par les Américains.

En Europe, les principaux serveurs sur le marché de l'information chimique sont STN, Lockheed, ESA et Télésystèmes, suivis par Data-Star, FIZ et SDC, ainsi que par Pergamon-Infoline, un mystérieux *outsider* chevauché par SDC. Derwent mis à part, les seules banques de chimie créées sur

place, réellement significatives, composent l'intéressante série de petites unités spécialisées dues à la Royal society of chemistry ; parmi celles-ci, les *Analytical abstracts* et les *Chemical business newsbase* apparaissent les plus consultées. Il est difficile de donner des chiffres précis, mais l'usage en Europe des banques de données sur la chimie semble représenter le quart de celui réalisé aux Etats-Unis.

Un autre facteur freinant l'utilisation des systèmes d'information en ligne en Europe tient aux coûts élevés des télécommunications et à leur manque de fiabilité. Une étude récente menée par EUSIDIC a fait apparaître que 30 % des échecs enregistrés dans les recherches venant d'Europe étaient dus aux défaillances des systèmes de télécommunications. Le problème de la langue est également un facteur d'explication. L'avenir pourrait bien voir la traduction automatique de tous les articles de recherche en une langue standard — probablement l'américain —, avant d'être chargés en ligne pour un accès international.

Politique et services en ligne

Derwent a toujours eu pour politique de faire payer un abonnement de base pour l'utilisation de ses différents services documentaires. Cet abonnement ne donne au client que le droit de choisir les documents, brochures, microfilms ou bandes magnétiques, produits par le service, et de décider de la forme et du volume des prestations. Comme il y a déjà eu un droit d'abonnement de payé, le tarif de fourniture des matériaux sélectionnés reste bas.

Cette formule est particulièrement intéressante pour les gros souscripteurs, qui ont besoin de nombreux exemplaires, et elle est également avantageuse pour Derwent, puisqu'elle garantit un engagement minimum de la part de l'abonné. C'est essentiellement pour cette raison que Derwent ne s'inquiète guère de l'impact que

pourront avoir les services en ligne sur ses ventes de produits conventionnels et qu'il les considère même comme un moyen de valoriser ses propres prestations aux yeux des utilisateurs, dans la mesure où ils requièrent moins d'efforts et une moindre immobilisation financière de leur part. Dans tous les cas de figure, Derwent ayant pour politique d'assumer l'ensemble des coûts de chargement sur le serveur et de contrôler ainsi les conditions d'accès et de prix, le principe de l'abonnement de base demeurait intact.

D'autres producteurs ont adopté une politique différente, élevant les prix de l'interrogation en ligne, mais cette mesure reste insuffisante. Ainsi les *Chemical abstracts* doivent-ils taxer à plusieurs centaines de dollars l'heure d'interrogation sur des serveurs extérieurs, pour compenser le déficit sur les ventes de produits papier. En outre, il est irréaliste de fixer des tarifs trop élevés en fonction du nombre de réponses obtenues, puisque ce sont précisément les recherches concernant des produits nouveaux, les plus précieuses et les plus importantes, qui n'obtiennent aucune réponse.

Il existe trois solutions rationnelles au problème qui se pose au producteur : faire payer, comme Derwent, un abonnement de départ ; restreindre les possibilités d'accès en ligne aux seuls clients ayant pris un abonnement pour des produits papier, ou faire payer, à l'exemple d'ISI, des tarifs très élevés aux non-abonnés ; devenir, tout comme les *Chemical abstracts*, son propre serveur.

Si les *Chemical abstracts* avaient opté pour cette politique dès le début, ils prêteraient moins le flanc à la critique. Dans l'état actuel des choses, après avoir laissé les différents serveurs investir un capital précieux en temps et en argent pour obtenir des fichiers CAS interrogeables, popularisant ainsi l'accès en ligne, ils leur ont fait une concurrence déloyale, retournant leur veste et affichant leur ambition de devenir la seule source d'information. En d'autres termes, les fichiers qu'ils fournissent aux autres serveurs sont moins complets, dans la mesure où ils ne comportent ni le texte des résumés ni les années antérieures.

En temps normal, un éditeur a le droit de protéger sa position en monopolisant les circuits de vente et de distribution de son produit. Le jeu habituel du marché protège le consommateur des dangers

R

d'une exploitation abusive. Cependant, avec la position d'exclusivité dont jouit la banque des *Chemical abstracts*, les mécanismes de protection du marché ne jouent pas et la communauté mondiale de la chimie traverse une situation de fortes turbulences.

Ou bien l'American chemical society accepte de donner les garanties voulues, fixant une politique tarifaire qui permette la protection de ses abonnés et assurant les serveurs contre tout risque de concurrence déloyale, ou bien, second terme de l'alternative, ses agissements peuvent donner lieu à des poursuites judiciaires intentées par les autres serveurs. Un organisme compétitif, financé par plusieurs pays, pourrait bien être considéré comme une émanation des communautés de chimistes, européenne et japonaise, afin d'assurer leur protection.

Les contraintes de caractère politique abondent dans le domaine de l'information sur les brevets, si importante pour les chimistes. Ainsi, l'Office américain des brevets, parallèlement aux offices des autres pays, transmet des bandes magnétiques à INPADOC, en Autriche, pour créer une bande de familles de brevets. Mais celle-ci n'est vraiment accessible, à un coût très élevé, que sur l'ordinateur INPADOC, installé à Vienne, ou en passant par Pergamon-Info-line à Londres.

Les enregistrements en ligne de l'Office européen des brevets ne sont publiquement accessibles à travers le monde qu'à partir d'un ordinateur installé en Europe et, même alors, par le canal d'INPADOC. Les résumés en anglais établis par l'Office japonais de brevets ne sont accessibles que sur SDC. De telles restrictions ne signifient pas seulement une menace au niveau commercial ; elles annulent l'avantage dû à la technique, qui permet d'accéder en même temps à plusieurs fichiers chaînés et d'effectuer des recherches croisées.

Le CD-ROM

Les CD-ROM sont les dernières vedettes apparues sur le marché de l'information chimique. On mène grand tapage autour d'elles, tant en Europe qu'aux Etats-Unis. Malgré l'abondance des études et des expositions qui leur sont consacrées en Europe, il n'existe que très peu de sociétés qui commercialisent les CD-ROM, aucune ne voulant d'ailleurs céder la

place aux autres. Quatre organismes se démènent pour promouvoir ce nouveau produit : Silver platter, BRS, Archetype et Harwell laboratory. On expérimente actuellement un certain nombre de systèmes partiellement financés par la Commission des Communautés européennes, dont le chargement sur CD-ROM de la base *Biotechnology abstracts* de Derwent. Il semblerait pourtant qu'à l'heure actuelle le CD-ROM soit la solution technique à un type de problème. Ce problème semble moins lié à l'accès en ligne qu'à la publication de travaux d'importance capitale, accessibles par mots clés et ne nécessitant pas de mise à jour trop fréquente. Ce qui inclurait pour les chimistes : *Dictionary of organic compounds*, *The Rubber handbook*, *Beilstein*, *Perry's chemical engineers handbook*, etc.

L'autre sphère évidente d'application pour le CD-ROM est la fourniture de documents. En Europe, le projet ADONIS (combinaison de Blackwell, Elsevier et Springer) renaît de ses cendres (cette fois sans Pergamon) pour voir si le CD-ROM fera l'affaire. Ses promoteurs ont décidé de concentrer leurs efforts sur les périodiques les plus demandés — les titres de biomédecine — et une expérience

de deux ans a été lancée, en collaboration avec plusieurs centres de fourniture de documents comme la British library, et avec le soutien de contrats passés avec l'Office européen des brevets pour la numérisation de 60 millions de pages de documents. L'équipement indispensable à chaque centre de fourniture de documents consiste en un micro-ordinateur, un lecteur de disques et une imprimante laser, le tout s'élevant à 12 000 \$ par poste de travail. Ce coût est au-delà des possibilités d'un service moyen d'information dans une entreprise, tout au moins dans l'industrie chimique.

De nouveaux services ?

Les inconvénients majeurs des systèmes actuels de recherche en ligne tiennent au volume des sorties-papier, aux besoins de sorties en différé et à la difficulté de repérer les résultats pertinents. En fait, essentiellement pour ces raisons, les recherches sont souvent menées sur une base étroite, inférieure à ce qu'elle devrait être, et seuls les titres sont imprimés, alors que le contenu intégral des résumés pourrait être fourni. En outre, les sorties imprimées ne peuvent reproduire de graphiques ou de formules chimiques, donnant ainsi

« L'utilisation des CD-ROM pour l'affichage devrait éliminer le plus gros du travail ingrat et quotidien d'une recherche, à savoir la chasse aux résumés pertinents avant de commander des copies du texte intégral via des systèmes tels qu'ADONIS »

des résultats sans signification. Pergamon a bien tenté de surmonter ce problème par l'emploi de disques optiques pour la recherche en ligne de brevets américains, en affichant sur écran séparé les premières pages des réponses, mais le projet a échoué pour des raisons prévisibles.

Derwent possède à présent un système suivant lequel un terminal intelligent peut rassembler des numéros d'accès issus d'une recherche en ligne, et retrouver directement la photographie des résumés à partir d'un microfilm contenant exactement 10 000 enregistrements par bobine de 60 mètres. Cependant, sauf dans le cas de DSI, le système présente l'inconvénient d'avoir à monter trop de bobines. Les systèmes de représentations graphiques pour entrer et retrouver des structures ont déjà été décrits. Même dans ce cas, la représentation graphique des réponses était loin d'être satisfaisante, surtout pour les structures de Markush dans les brevets, en raison de la surcharge de l'écran.

L'emploi du CD-ROM comme substitut aux énormes possibilités de recherche des systèmes com-

merciaux en ligne, surtout lorsque ceux-ci s'appuient sur d'énormes banques de données, n'étant pas prévu, particulièrement dans le cas d'une recherche croisée par fichiers, son avantage immédiat tiendrait à sa capacité de stockage de 600 megabytes; l'application porterait sur le stockage de résumés comportant des graphiques et des formules chimiques stockés à des fins d'affichage seulement pour de grosses banques très consultées. Les mises à jour périodiques peuvent se faire une fois par an seulement, l'information la plus récente étant conservée entre-temps sur microfilm ou sur carte laser.

Un tel système appliqué au *Chemical patents index* de Derwent, par exemple, permettrait de récupérer un des bénéfices les plus appréciables du service, disparu en même temps que les cartes perforées. Au départ, les recherches s'effectuaient grâce à des cartes perforées comportant des résumés, les résultats pouvant alors être appréhendés dans toute leur splendeur — schémas, formules et texte en clair. L'utilisation des CD-ROM pour l'affichage de-

vrait éliminer le plus gros du travail ingrat et quotidien d'une recherche, à savoir la chasse aux résumés pertinents avant de commander des copies du texte intégral *via*, si tout va bien, des systèmes tels qu'ADONIS.

Il apparaît donc que plusieurs secteurs du marché international pour l'information chimique appartiennent aux Européens en général, et à Derwent en particulier. Derwent qui, en réalité, tire un tiers de ses revenus annuels de 30 millions de dollars des Etats-Unis et un autre tiers du Japon. Ses revenus représentent la moitié de ceux des puissants *Chemical abstracts*, et ses 60 000 heures annuelles de consultation en ligne rangent ses fichiers parmi les plus utilisés. Si l'on ajoute à cela la position unique de la banque de données INSPEC sur l'électronique et la physique, l'Europe ne se défend pas si mal sur la place du marché de l'information scientifique.

Un seul os : Derwent appartient à présent à une compagnie canadienne dont le siège social se trouve à New York.