
BIBLIOGRAPHIE SPÉCIALISÉE

LES coûts de production de l'information bibliographique font intervenir des coûts d'acquisition des documents, de traitement, catalogage et indexation, et de fabrication proprement dite. Dans les deux cas développés ici, l'achat des documents ne pèse que partiellement ou pas du tout sur les coûts de la banque de données. Au CDST, les périodiques servent à la fois à la bibliothèque, à la fourniture de documents primaires et à PASCAL et les frais d'abonnement sont donc répartis entre ces trois services.

Pour la *Bibliographie de l'histoire de France*, la Bibliothèque nationale offre la possibilité de travailler sur ses collections de périodiques. C'est donc essentiellement des autres coûts, traitement et fabrication, qu'il est question.

La production d'information en ligne n'est pas dissociable de celle de ses dérivés papier ou autres. Pour PASCAL, une seule saisie de l'information génère plusieurs produits de diffusion. La *Bibliographie de l'histoire de France* part d'une bande de

photocomposition, là encore pré-traitée, pour fabriquer la banque MARIANNE. Un même imprimeur, Jouve, est à l'origine de cette saisie à plusieurs usages. Elle permet de dépasser le dilemme, mettre en banque de données un ouvrage en transcodant la bande de photocomposition, exercice laborieux qui peut se révéler plus compliqué que la ressaisie totale, ou bien éditer une banque de données, avec le risque d'obtenir un résultat bien pauvre sur le plan typographique.

PASCAL en chiffres

par

et

Odile Artur

Service technique du CDST

Denise Pelissier

Service diffusion du CDST

Le Centre de documentation scientifique et technique (CDST) du CNRS a été créé en 1939, initialement pour les besoins propres du CNRS. Il s'est ouvert progressivement à une communauté scientifique et technique de plus en plus large : chercheurs, industriels et ingénieurs du monde entier.

A vocation multidisciplinaire, le CDST collecte, analyse, indexe et diffuse la plus grande partie de la littérature mondiale dans tous les domaines scientifiques et techniques : sciences exactes, sciences de la vie, sciences de la terre et technologie. Pour faciliter l'accès à la littérature mondiale, le CDST a mis en place des bases de données bibliographiques souvent en coopération avec des organismes français et étrangers.

Les bases de données bibliographiques

Elles sont au nombre de trois : PASCAL, la plus importante, IALINE et WORLD TRANSINDEX.

PASCAL

C'est une base multidisciplinaire qui permet de glaner sur un sujet précis des informations émanant de disciplines diverses et dispersées dans de multiples sources. Cette base de données a démarré avec la création du centre en 1940. Elle est automatisée depuis 1973 et comporte à ce jour 6 millions de références bibliographiques accessibles directement en ligne. Pour répondre aux impératifs des sciences et des techniques d'aujourd'hui, multidisciplinarité et spécialisation, la base PASCAL est organisée en deux types de fichiers : PASCAL M et PASCAL S.

PASCAL M couvre l'essentiel de la littérature mondiale dans les disciplines fondamentales de la physique et de la chimie, les sciences de la vie, les sciences appliquées et les technologies, les sciences de la terre et les sciences de l'information. Elle s'accroît de 300 000 références par an. Plus de 4 500 périodiques sont dépouillés, pro-

venant de 107 pays (Europe, pour 57 %, Etats-Unis, 33,6 %, autres, 9,4 %). Des comptes rendus de congrès, des thèses, des rapports et des ouvrages sont également analysés.

PASCAL S est un ensemble de bases sectorielles qui répondent au besoin d'exhaustivité de certains domaines dans lesquels « l'éparpillement » des sources nécessite un traitement plus approfondi. Ces bases sectorielles sont actuellement au nombre de 10, parmi lesquelles on peut citer, par exemple, les sciences de l'information, l'énergie, les biotechnologies et la médecine tropicale.

Dans chacune des bases de PASCAL S, la couverture est aussi complète que possible. En plus des périodiques spécialisés de chaque domaine (4 000 titres environ), un grand nombre de publications sont examinées avec un effort important sur la littérature grise. Les documents analysés sont pour 63 % d'entre eux en anglais, 12 % en français, 10 % en

russe et 8% en allemand. Dans PASCAL, chaque signalement comporte différents éléments décrivant le contenu scientifique de l'article : titre, résumé, mots clés, code de classement, et le catalogue : auteurs, affiliation, référence bibliographique.

IALINE

Cette base de données concerne les industries agro-alimentaires. Elle est coproduite par le CDST et le CDIUPA (Centre de documentation des industries utilisatrices de produits agricoles). Cette coproduction a démarré en 1982. La base s'accroît annuellement de 13 000 signalements.

WORLD TRANSINDEX

Unique en son genre, cette base de données signale les traductions de documents dans des langues difficiles (slaves, asiatiques, japonaise, etc.) vers les langues occidentales (anglais, français, allemand, espagnol), ainsi que les traductions interoccidentales.

Créée en 1978, par fusion des bibliographies suivantes : *Index* du Centre international de traduction (CIT), *Translation bulletin* de la CEE et le *Bulletin des traductions* du CDST, elle est coproduite par le CIT et le CDST. Depuis janvier 1987, le National translation center (NTC) envoie sur bande magnétique l'ensemble des traductions réalisées aux Etats-Unis ; la base de données s'appelle désormais WORLD TRANSLATIONS INDEX.

Contrairement aux deux bases de données décrites précédemment, chaque signalement comporte un double catalogage : la description du document traduit et la description du document source.

Fabrication de la base PASCAL

Volumes traités

Ils représentent 1 800 documents par jour, soit 9 000 par semaine et 380 000 par an, auxquels il faut ajouter 40 000 documents fournis par certains coopérateurs sur support magnétique (bandes, disquettes).

Principales étapes de la fabrication

- le bordereau : catalogage et analyse scientifique ;
- la saisie : sous masque (catalogage) ; au kilomètre ;
- le premier contrôle ;
- la lecture et les corrections ;

- le deuxième contrôle et la validation ;
- la génération.

Une bande de 9 000 documents, en format PASCAL BASE, est constituée toutes les semaines. Le catalogage des articles de périodiques est assuré selon deux circuits : le remplissage manuel des bordereaux, par une dizaine de collaborateurs extérieurs pigistes (800 par jour) ; la saisie directe sur écran, sous masque (1 000 docu-

ments par jour), prise en charge par les opérateurs de la société Jouve. Ce deuxième circuit, mis en place en 1983, représente une nette amélioration : suppression de l'écriture sur bordereau ; optimisation de la saisie avec génération des constantes (les niveaux périodique et fascicule ne sont saisis qu'une seule fois pour x analyses) ; mise à disposition rapide des données stockées (pour les revues de sommaires par exemple). L'analyse scientifique

des documents est écrite sur bordereau par les rédacteurs. Elle est saisie au kilomètre, puis mixée aux données de catalogage préalablement saisies.

Tout document pris en compte est contrôlé : statut des zones (obligatoires, facultatives, aléatoires) ; contrôles syntaxiques et sémantiques ; contrôle par rapport à des fichiers d'autorité dont un lexique de 100 000 termes et un fichier de périodiques de 12 000 titres.

Tout document conforme après contrôle passe ensuite dans une phase de génération. Avec l'accès au lexique : génération des mots clés anglais, espagnols, allemands, et des termes autopostés. Avec l'accès au fichier des périodiques : génération des titres abrégés et complets des périodiques. Avec l'accès au fichier des congrès : génération des vedettes congrès.

Le document saisi comporte en moyenne 700 signes, le document généré 1 000 à 1 300 signes. Aux économies de saisie et de lecture s'ajoutent la fiabilité et la cohérence des données générées.

Développement d'un logiciel de saisie

Nous développons actuellement en collaboration avec Jouve un logiciel de saisie d'informations logiques PSLOG sur micro MS-DOS, qui possédera, entre autres, les avantages suivants :

— Editeur de documents structurés, il permettra des gains de saisie par : génération des valeurs fixes pour une étape de travail ; génération de valeurs d'un document sur un autre document par mémorisation indiquée par l'utilisateur ; suppression de la saisie des éléments définissant la structure.

Il permettra des contrôles, dès la saisie, du contenu des zones et de la cohérence du document.

— Editeur multifenêtre, il offrira :

Sur le plan de la visualisation, une personnalisation de la vue de l'information ; un jeu étendu de caractères (langues d'origine latine, caractères accentués, grec, symboles mathématiques).

Sur le plan de l'ergonomie, le déplacement simple dans le document par touches liées à la visualisation (changement de fenêtres, changement d'écrans), ou liées à la structure d'un document (ex : occurrence suivante) ; le défilement horizontal ou vertical de

toute l'information dans une fenêtre (zones de longueur variable dans une fenêtre de dimensions fixes) ; la fonction de base d'un traitement de texte : saisie, suppression, modification, déplacement, copie.

L'insertion de ce logiciel dans la chaîne de fabrication devrait supprimer peu à peu l'écriture sur bordereau ; diversifier et rendre plus autonome les pôles et les modes d'entrée des documents ; diminuer les flux inutiles d'information ; réduire les délais de fabrication ; optimiser les coûts de constitution de la base.

Diffusion des différents produits

Des produits sont disponibles sur différents supports : papier, microfiche, bande magnétique.

Publications bibliographiques

Au nombre de 79, elles se présentent en quatre collections : *Pascal sigma*, qui est la version imprimée de PASCAL M ; *Pascal thema*, réalisée à partir des bases sectorielles PASCAL S ; *Pascal folio* qui est constituée de tirés à part des deux précédentes ; et *Pascal explore*, qui permet sur un sujet pluridisciplinaire, de regrouper l'ensemble des informations de la base PASCAL. Elles sont disponibles sous forme d'édition papier ou de microfiches.

Profils documentaires

Ce sont des bibliographies mensuelles signalant, sous forme de références bibliographiques, les publications récentes de textes et d'articles scientifiques se rapportant à un thème précis.

Il en existe deux catégories :

— les profils standard, dont les sujets sont déterminés au CDST par les spécialistes des domaines. Environ 300 titres sont proposés dans tous les domaines scientifiques et techniques. L'abonnement, pour 15 livraisons, est annuel.

— les profils personnalisés, dont les sujets sont proposés par les utilisateurs eux-mêmes.

Bandes magnétiques

PASCAL peut être fourni en totalité ou en partie pour exploitation (recherche documentaire), sur tout ordinateur disposant d'un logiciel documentaire et quel que soit le logiciel.

Recherches bibliographiques rétrospectives

Les recherches bibliographiques rétrospectives permettent de faire le point sur une question donnée depuis 1973. Elles sont réalisées au CDST sur demande, sur tout ou partie du fichier, par des spécialistes.

Interrogation en ligne

La base PASCAL est accessible en totalité sur deux serveurs : QUESTEL-TELESYSTEMES (logiciel QUESTEL PLUS) ; ESA-IRS (logiciel QUEST). Elle est en cours d'implantation sur DIALOG. Le SUNIST reçoit les thèses de sciences provenant de PASCAL, pour alimenter TELETHESES.

Accès aux documents originaux

Tous les documents repérés dans PASCAL peuvent être obtenus sous forme de photocopies ou de microfiches par l'intermédiaire du CDST et ce, par courrier, par télex ou directement en ligne. La commande en ligne est disponible sur tous les serveurs.

Les utilisateurs de PASCAL

Ils sont très divers, bibliothèques, centres de documentation, individus, et répartis dans le monde entier. Il y a 9 500 abonnés aux publications sur papier (dont presque la moitié pour l'étranger).

Le service des profils documentaires est en baisse régulière. De 3 400 abonnements en 1979, il est passé à environ 1 000, en raison principalement de la concurrence de l'interrogation en ligne. Près de 90 % des abonnés sont en France. La diffusion sur disquette démarre et compte pour le moment 37 abonnements. Les bandes magnétiques sont diffusées aux coopérateurs et coproducteurs du CDST, et aux serveurs ; à l'étranger, au Japon, en Pologne, en Suède et au Brésil.

Plus de 500 recherches bibliographiques rétrospectives sont effectuées chaque année. Plus de 80 % des demandes concernent le secteur bio-médical. Les clients sont les laboratoires pharmaceutiques, des sociétés privées, des laboratoires de l'université et du CNRS. 80 % des demandes émanent de France.

PASCAL est utilisé, en ligne, dans plus de 30 pays différents. La répartition pour les principaux pays sur les deux serveurs est la sui-

vante : sur Questel, 90 % pour la France, 2 % pour l'Europe de l'Ouest, 3 % pour l'Amérique du Nord, 5 %, autres. Sur ESA, 61 % pour la France, 31 % pour l'Europe de l'Ouest, 0,5 % pour l'Amérique du Nord, 7,5 %, autres. Au total, en 1985, il y a eu 11 334 heures d'interrogation, 1 350 000 signalements édités, visualisés ou télédéchargés.

Coûts de production et de diffusion

Le CDST utilise une comptabilité analytique qui permet de répartir les coûts suivant les produits. Les coûts de fabrication sont bien connus, par contre les coûts de diffusion de certains produits (bandes magnétiques, profils documentaires, recherches bibliographiques) sont plus difficiles à maîtriser. Les coûts de saisie varient suivant le mode d'alimentation. Dans le cas de la saisie au kilomètre de bordereaux, le coût est de 32,51 F par mille de caractères ; pour une bande magnétique provenant d'un coopérant, les coûts de prise en compte comprennent un coût forfaitaire de 227,39 F par bande et un coût de 2,43 F par millier de caractères traités.

Depuis janvier 1985, le CDST a expérimenté la saisie décentralisée sur micro-ordinateur ; il est possible d'alimenter la chaîne par disquette ou par télétransmission.

Globalement, les coûts de fabrication des bases de données du CDST sont inférieurs à ceux d'autres organismes. Ceci est dû au grand nombre de documents traités et à l'automatisation de plus en plus poussée de la chaîne de production. Le développement de fichiers d'autorité a permis de réaliser d'importantes économies de saisie, tout en améliorant la qualité des bases, et de diversifier les produits et les services dérivés des bases.

Coûts de constitution d'une notice PASCAL et d'une notice WTI

	PASCAL	WTI
acquisition	12,44 F	0
catalogage	4,34 F	13,02 F
traitement conceptuel (analyse, indexation)	67,41 F	30,52 F
gestion technique des bordereaux	7,79 F	12,79 F
saisie (traitement informatique)	29,51 F	17,00 F
total entrée	109,05 F	73,33 F
total (y compris acquisition)	121,49 F	73,33 F

Les différences de coût unitaire proviennent des différences de traitement. Pour WTI, il n'y a pas d'analyse, l'indexation est minimale : deux ou trois macro-descripteurs par bordereau et un code de classement réduit ; en revanche le catalogage est plus complexe : double catalogage et nombreuses vérifications manuelles des éléments.

Coûts d'édition et de diffusion des bulletins bibliographiques

	PASCAL	WTI
coûts CDST (coordination, suivi)	2,52 F	2,52 F
traitement informatique et coûts d'impression et de routage	8,48 F	8,43 F
papier	0,50 F	0,83 F
total édition-diffusion	11,50 F	11,78 F

Coûts moyens des profils documentaires

formation, contrôle, maintenance	180 F/profil/an
traitement informatique (15 mises à jour par an)	373 F/profil/an
total fabrication	555 F/profil/an

Il faut souligner que les documents primaires qui alimentent la base de données PASCAL servent également au service de fourniture de documents et à la bibliothèque. Les coûts d'acquisition de ces documents sont ventilés à égalité pour ces trois utilisations.