

**PASCAL : UNE BASE DE DONNÉES
MULTIDISCIPLINAIRE SON UTILISATION EN
PHYSIQUE ATOMIQUE ET MOLÉCULAIRE ET
PHYSIQUE DES FLUIDES ET DES PLASMAS**

J. Buhr, C. Degen

► **To cite this version:**

J. Buhr, C. Degen. PASCAL : UNE BASE DE DONNÉES MULTIDISCIPLINAIRE SON UTILISATION EN PHYSIQUE ATOMIQUE ET MOLÉCULAIRE ET PHYSIQUE DES FLUIDES ET DES PLASMAS. Journal de Physique Colloques, 1977, 38 (C3), pp.C3-249-C3-251. 10.1051/jphyscol:1977327 . jpa-00217115

HAL Id: jpa-00217115

<https://hal.archives-ouvertes.fr/jpa-00217115>

Submitted on 1 Jan 1977

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PASCAL : UNE BASE DE DONNÉES MULTIDISCIPLINAIRE SON UTILISATION EN PHYSIQUE ATOMIQUE ET MOLÉCULAIRE ET PHYSIQUE DES FLUIDES ET DES PLASMAS

J. M. BUHR et C. DEGEN

Centre de Documentation du C.N.R.S., 26, rue Boyer, 75971 Paris Cedex 20, France

Résumé. — Nous décrivons le système PASCAL du Centre de Documentation du C.N.R.S., base de données multidisciplinaires. PASCAL est un système automatisé de saisie, traitement et diffusion sélective d'informations bibliographiques dans de larges domaines scientifiques et techniques. Le système PASCAL fournit des bandes magnétiques, des Bulletins signalétiques, des profils documentaires et des recherches rétrospectives tant par traitement par lots qu'en mode conversationnel. Pour l'illustrer nous donnons un exemple en Physique atomique et moléculaire.

Abstract. — We describe the system PASCAL of the « Centre de Documentation » of C.N.R.S., which deals with a multidisciplinary data base. PASCAL is an automated system for input, treatment and selective dissemination on a wide scope of scientific and technical fields. Its products are tape series, « Bulletins Signalétiques », documentary profiles, retrospective searching as well in batch as on line. As illustration, we give an example in atomic and molecular Physics.

INFORMASCIENCE, Centre de Documentation du C.N.R.S., utilise le système informatique PASCAL (Programme Appliqué à la Sélection et la Compilation Automatique de la Littérature) pour la saisie, le traitement et la diffusion d'environ 500 000 références bibliographiques chaque année qui s'ajoutent aux 8 000 000 de références existantes.

1. **La base de données PASCAL.** — Cette base de données bibliographiques est essentiellement encyclopédique en matière d'information scientifique et technique. Elle recouvre l'ensemble des Sciences fondamentales, des Sciences de la Terre, des Sciences de l'Ingénieur, des Sciences Biologiques et Médicales et des Sciences Agricoles.

Les sources de ce fichier documentaire sont constituées en majeure partie par des revues scientifiques et techniques primaires (13 000 titres reçus régulièrement), mais également par des thèses (55 000 en stock), des comptes rendus de Congrès (17 000 environ), des rapports de recherche et développement, notamment ceux de la Délégation Générale à la Recherche Scientifique et Technique (3 500 répertoires) et des ouvrages, principalement des livres édités ou subventionnés par le C.N.R.S. et un fonds spécial important d'ouvrages russes.

Les articles scientifiques et techniques contenus dans ces divers types de documents sont identifiés à la fois :

— par un catalogage très précis obéissant aux règles internationales élaborées par l'UNISIST-ICSU-AB, catalogage faisant apparaître un certain nombre de

paramètres (auteurs, adresse d'auteur, titre, type de document, langue, etc.), qui peuvent lors d'un traitement constituer autant de clés de tri ;

— par une analyse du contenu se présentant sous la forme d'un résumé signalétique décrivant brièvement les éléments essentiels du document ;

— par un code de classement permettant ainsi d'organiser le fichier documentaire en regroupant les signalements relatifs à un même sujet ;

— par une indexation à l'aide de descripteurs contrôlés, d'identificateurs et de mots-clés libres permettant une recherche automatique, fiable et économique, de références après saisie.

2. **Produits fournis par le système PASCAL.** — Tous les documents catalogués, analysés et indexés, sont saisis sur bande magnétique. Cette bande, après avoir subi certains traitements, est utilisée pour fournir divers types de produits documentaires :

2.1 **LES BANDES MAGNÉTIQUES.** — Elles constituent le premier produit du fichier PASCAL. Leur contenu global est celui de la base de données organisée en macrodomaines ou unités documentaires. Leur diffusion peut s'effectuer suivant le centre d'intérêt de l'utilisateur (unité documentaire, section du Bulletin signalétique ou profil sur bande magnétique).

2.2 **LE BULLETIN SIGNALÉTIQUE.** — Chaque mois, l'ensemble des signalements saisis est ventilé grâce au code de classement dans une ou plusieurs des 49 sec-

tions éditées par photocomposition, chaque section correspondant à un domaine scientifique ou technique (exemples : Section 120, Astronomie. Physique spatiale. Géophysique ; Section 145, Electronique). Chaque bulletin se présente donc comme l'ensemble de références relatives à un domaine, rangées selon un plan de classement élaboré en liaison étroite avec des scientifiques spécialistes du domaine. Parfois, comme c'est le cas en Physique, ce plan de classement est commun à divers Centres de Documentation français et étrangers (Bulletin signalétique, American Institute of Physics, Physikalische Berichte, INSPEC-Physics Abstracts). Des index auteurs et matières complètent chaque bulletin et facilitent l'accès à l'information.

Ainsi le Bulletin signalétique Section 165 : Physique atomique et moléculaire, Physique des fluides et des plasmas, a porté à la connaissance des Scientifiques environ 17 000 références d'articles, dont 14 000 relatifs aux atomes et molécules et 3 000 relatifs aux fluides et plasmas en 1975.

Des index hiérarchisés ou permutés (ces derniers engendrés automatiquement par ordinateur) sont constitués à partir de mots-clés choisis parmi les termes d'indexation, issus d'un lexique contrôlé, et font apparaître les facettes essentielles des signalements publiés : matériaux (ex. : hydrogène atome, hydrogène ion moléculaire, plasma hors équilibre), phénomènes ou processus (ex. : moment dipolaire, collision atome-ion, excitation, échange de charge, instabilité paramétrique), conditions opératoires (ex. : basse température, haute pression, champ haute fréquence, moyenne énergie), méthodes de calcul (ex. : approximation de Glauber) ou techniques expérimentales (ex. : spectrométrie de fluorescence).

2.3 LA DIFFUSION SÉLECTIVE DE L'INFORMATION (profils documentaires). — Cette diffusion permet de fournir à l'utilisateur de la documentation « à la carte ». Elle évite au chercheur ou à l'ingénieur d'avoir à consulter divers bulletins bibliographiques en lui donnant tous les mois les signalements d'articles les plus récents correspondant à son centre d'intérêt particulier. Il existe deux types de profils :

— les profils standards, relatifs à des problèmes classiques ou à des techniques de pointe, dont les listes sont régulièrement diffusées auprès des laboratoires et des entreprises. Ce type de bibliographie peut être considéré comme des microbulletins signalétiques et être diffusé à plusieurs utilisateurs (exemples : polarisabilités électriques : théorie et expériences ; études théoriques et expérimentales des transitions électroniques d'ions atomiques ; Tokamak ; instabilité paramétrique) ;

— les profils personnalisés répondant à une question particulière d'un seul demandeur, les sujets de ces profils n'étant pas révélés. Ce pourrait être par exemple : ionisation en couche interne et émission

de rayons X d'éléments de la première série de transition, soumis à un impact d'ions de haute énergie.

Le caractère multidisciplinaire de la base de données permet de répondre à des questions de tout type et relatives à des domaines intersectoriels. Ainsi des données sur les atomes, les molécules ou les plasmas concernent aussi bien des astrophysiciens ou des chercheurs étudiant les lasers.

2.4 LES RECHERCHES BIBLIOGRAPHIQUES RÉTROSPECTIVES. — Alors que la diffusion sélective de l'information fournit les références les plus récentes entrées dans le fichier, la recherche rétrospective permet la diffusion des signalements correspondant à une question particulière de documents publiés dans les années antérieures. Le fichier PASCAL a commencé à être automatisé en 1971 et des recherches peuvent donc être entreprises depuis cette date à l'aide de l'ordinateur ; pour les années antérieures (1940-1970), elles sont éventuellement complétées manuellement.

Le programme VIRA est utilisé lors de la recherche documentaire pour la diffusion sélective de l'information et la recherche rétrospective en traitement par lots. Il est fondé sur l'exploitation directe du langage naturel (chaîne de caractères). Chaque question renferme trois types de données :

- les paramètres d'identification du profil,
- les chaînes de caractères opérantes (descripteurs, mots-clés libres, etc.) rassemblés en groupe. Des indications concernant les zones du bordereau PASCAL (titres, auteurs, indexation, etc.) sur lesquelles doit être faite la recherche les accompagnent, ainsi qu'une pondération de chaque notion permettant de classer par ordre de pertinence décroissante les références répondant à la question posée au fichier,
- une équation logique combinant les groupes d'opérantes avec divers opérateurs (booléens, arithmétiques, de troncature).

Pratiquement, tant pour les profils documentaires que pour les recherches rétrospectives, l'interrogation est effectuée principalement sur la zone comprenant les termes d'indexation contrôlés ; elle peut être complétée ou affinée par une recherche sur le code de classement, sur le titre ou le résumé ou sur la zone des mots libres. Cette dernière zone constitue une source d'information complémentaire lorsque se développent de nouvelles théories ou de nouvelles techniques expérimentales, ne figurant pas encore dans le lexique de termes contrôlés (ex. : spectroscopie sans effet Doppler).

Alors commence un dialogue de plusieurs mois entre le demandeur et le rédacteur (ou les rédacteurs) spécialiste du domaine, afin de mieux comprendre le sens que chacun d'eux attribue aux mots, d'abord pour la mise au point de l'écriture du profil, puis, après les trois premières sorties du profil documentaire, pour modifier, élargir ou restreindre l'interrogation suivant les réponses obtenues.

Exemple de profil personnalisé avec sa formulation :
Etude théorique de l'excitation rotationnelle de molécules diatomiques minérales par impact. Etudes expérimentales sur l'azote.

Ainsi pour le profil donné en exemple, le demandeur est surtout intéressé par l'excitation par impact d'atomes, éventuellement par l'excitation par impact d'ions ou de molécules et les études sur des composés organiques doivent être totalement exclues.

La première démarche sera de chercher les termes significatifs de la question principalement dans le lexique de termes contrôlés de la Section 165 A (Physique atomique et moléculaire), puis de les ranger en groupes (ex. : groupe A : excitation ; groupe B : état rotationnel excité + état rovibrationnel excité + état rovibronique excité ; groupe C : collision atome-molécule + collision ion-molécule + collision molécule-molécule ; groupe D : étude théorique ; groupe E : étude expérimentale). Le programme VIRA permet d'interroger jusqu'à 26 groupes pouvant contenir jusqu'à 128 termes comportant chacun au maximum 34 caractères. Certains termes ne figurant pas dans le lexique contrôlé (candidats descripteurs, identificateurs de matériaux [ex. : azote molécule], mots-clés libres) peuvent également entrer dans ces groupes. C'est alors que l'on précise la zone où doit s'effectuer la recherche : pour le code d'orientation (ex. : 165 A), les mots contrôlés, les identificateurs, les mots-clés libres. Des recherches pourraient aussi être faites sur le titre, sa traduction et le résumé signalétique, sur la cote du périodique, la référence bibliographique, les auteurs, le code langue, etc. On peut aussi rechercher un terme seul (ex. : excitation) en excluant les descripteurs composés qui le comprennent (excitation échelonnée, transfert énergie-excitation) et qui sortiraient normalement avec une recherche classique.

La logique de recherche se poursuit en pondérant les termes de chaque groupe. Dans l'exemple choisi, les articles traitant de l'excitation rotationnelle par collision atome-molécule seront les plus pertinents (on affecte à ce terme un poids + 08, supérieur aux poids + 02 de collision ion-molécule ou collision

molécule-molécule) ; l'état rotationnel excité pur (poids + 06) intéresse davantage le demandeur que les états rovibrationnels ou rovibroniques excités (poids + 04 ou + 02).

Il ne reste plus qu'à écrire l'équation logique combinant groupes opérands et opérateurs. Exemples : (*) pour *et*, + pour *ou*, > pour exclusion totale. D'autres opérateurs plus subtils permettent soit le classement des groupes par ordre d'intérêt, soit l'exclusion relative d'un groupe de moindre intérêt, soit l'exclusion relative d'un terme à l'intérieur d'un groupe. Enfin une commande $Z = x$ donne la possibilité de limiter le nombre x de documents sortis, évidemment les plus pertinents du fait de la pondération.

2.5 PASCALINE. — PASCALINE ou PASCAL on LINE a été mis en place pour améliorer la rapidité et l'efficacité de tous les services, car le mode conversationnel permet :

- la simplification de toutes les procédures de recherche,
- la souplesse d'emploi,
- la décentralisation au niveau des utilisateurs.

La mise en place de ce nouveau service nécessite un logiciel documentaire, un ordinateur support de ce logiciel et un réseau de distribution. Actuellement PASCALINE est opérationnel :

- sur le réseau Cyclades avec le logiciel MISTRAL sur les ordinateurs IRIS 80 à Lyon et Grenoble,
- sur le réseau ESRO, raccordé lui aussi au réseau Cyclades, avec le logiciel RECON sur ordinateur IBM 360-50 basé à Frascati (Italie).

La mise en place du réseau de télécommunications européen EURONET qui sera opérationnel à la fin de l'année 1977 permettra d'améliorer et d'étendre encore les services.

Ainsi la documentation automatisée est devenue une réalité. Il reste à souhaiter que les demandeurs, chercheurs scientifiques et techniques, toujours plus nombreux, fassent connaître et leurs besoins et leurs critiques afin de pouvoir être mieux servis, aidant ainsi les documentalistes dans leurs tâches.