
Bulletin de l'Union des Physiciens

Association des professeurs de Physique et de Chimie

Éditorial

De la table quinquennale...

...à la base de données

par Monique SCHWOB
Rédacteur en chef

Cette huitième table quinquennale, que beaucoup d'entre vous attendent avec impatience, fait suite à celles dressées successivement par M. SAVARIAU, Mlle COURTIN, M. LETELLIER, M. LE REVEREND et M. CHARPENTIER pour les trois dernières. Ce n'est qu'à partir de la sixième édition que, devant l'augmentation du volume des bulletins (déjà...), cette publication devient «quinquennale». Malgré le retard avec lequel paraît cette huitième édition, nous avons pensé qu'il était souhaitable de conserver la même périodicité, c'est pourquoi c'est une table quinquennale (1991-1995) que nous vous proposons aujourd'hui dans une version «papier» ; la version informatisée, disponible à la rentrée intégrera (en principe...) l'année 1996.

LE GROUPE DE TRAVAIL «BASES DE DONNÉES»

Devant l'évolution actuelle des nouvelles technologies de l'information et de la communication (NTIC), l'informatisation croissante des Centres de Documentation des établissements scolaires et des laboratoires de physique et de chimie, mais surtout devant le nombre croissant d'enseignants disposant à domicile d'un ordinateur et bientôt de systèmes multimédia et d'accès Internet, il nous a semblé nécessaire de réfléchir à l'utilisation de ces nouvelles technologies dans le cadre du BUP.

Nous avons donc, il y a un peu plus d'un an, constitué un groupe de travail «base de données» qui s'est donné pour objectif de réfléchir à la valorisation du potentiel du BUP en utilisant, entre autres, les nouvelles technologies. Ce groupe s'est constitué à partir de quelques collègues déjà intéressés par ce sujet et de «nouveaux venus» ayant répondu à un appel paru dans le BUP (juin 1995).

Qu'il me soit tout de suite permis de remercier tous les membres de ce groupe* qui ont consacré depuis plus d'un an (au cours de nombreuses réunions et d'encore plus nombreuses journées de travail à domicile...) un temps important à une réflexion indispensable, puis à un énorme travail d'indexation des six dernières années du BUP, et enfin à une mise en commun et à une homogénéisation de l'ensemble des travaux. Parallèlement deux membres du groupe (Olivier KEMPF et Claude LACOMBE) se sont penchés sur les problèmes de conception et d'élaboration d'une base de données automatisée et des outils indispensables à son exploitation. Cette base de données sera disponible sur disquette dès la rentrée scolaire 1997-1998 et sur Internet par l'intermédiaire du serveur UdP dès que celui-ci sera ouvert. Nous en reparlerons.

Je profite de cette occasion pour remercier à nouveau J.-P. CHARPENTIER, auteur des trois dernières tables des matières : nous mesurons mieux aujourd'hui l'immensité de la tâche accomplie, seul et sans les moyens techniques dont nous disposons actuellement ! Qu'il soit d'autant plus remercié que nous sommes partis de son travail et d'un certain nombre des idées qu'il avait développées, en particulier de la liste de mots qu'il avait élaborée, pour amorcer notre réflexion.

Le travail de ce groupe ne devrait pas s'arrêter avec la publication de cette table sous forme papier/disquette/Internet. Après avoir exposé brièvement les idées et les principes retenus pour cette première étape, nous évoquerons les projets à venir.

TABLE ET INDEX «PAPIER» ET BASE DE DONNÉES INFORMATISÉE

Nous nous sommes donné comme premier objectif de réaliser simultanément la table quinquennale dans sa version «papier» et une version informatisée disponible sur disquette et sur le serveur UdP en cours d'élaboration sous la responsabilité de la commission informatique.

Deux questions majeures se sont alors posées, l'une concernant l'aspect documentaire, l'autre l'aspect informatique. Sur le plan documentaire, la question était celle de

* J. CARRETTO, O. CARRIER, P. GOUVERNEUR, O. KEMPF, C. LACOMBE, B. MONTFORT, G. MORA, A. NOËL, J.-F. PIGNAL, M. PROFIT, M. SCHWOB, F. VILLEDARY.

l'outil d'indexation à utiliser, thesaurus ou non et si oui, lequel ? Sur le plan informatique il s'agissait de savoir si nous décidions d'utiliser un logiciel commercial ou si nous allions créer nos propres outils. Les deux questions étaient évidemment partiellement liées dans la mesure où le choix d'un thesaurus classique (Motbis couramment utilisé dans les CDI par exemple) nous aurait conduits à utiliser le logiciel de recherche documentaire correspondant.

Thesaurus ou non ?

L'un des principaux problèmes de l'accès à toute documentation est la cohérence entre l'indexation et la recherche : les mots utilisés pour répertorier un article, un ouvrage, ou tout autre source documentaire, doivent être les mêmes que ceux susceptibles d'être formulés par l'utilisateur. *L'indexation ne doit donc pas décrire un article mais servir à le retrouver ce qui implique immédiatement qu'il n'y a pas d'indexation indépendante de l'utilisateur potentiel.* Ce point fondamental a orienté nos choix.

Un *thesaurus* est un ensemble structuré de mots clés qui permet de rationaliser à la fois l'indexation et la recherche dans la mesure où l'outil est commun à l'indexeur et à l'utilisateur. C'est une liste de mots clés définie à l'avance et structurée par un certain nombre de relations entre les mots (hiérarchie, association, proximité, ...). La qualité d'un thesaurus dépend d'une part de son organisation propre, d'autre part, et surtout, de son adéquation par rapport aux besoins des utilisateurs et de sa capacité à évoluer pour intégrer de nouveaux concepts au fur et à mesure des besoins.

Le thesaurus MOTBIS des CDI de lycées

Beaucoup de CDI de lycées utilisent le thesaurus MOTBIS diffusé par le CNDP (et le logiciel BCDI diffusé par le CRDP de Poitiers qui indexe également un certain nombre de revues dont on peut recevoir, par abonnement, les fiches de dépouillement).

Nous avons étudié ce thesaurus pour voir dans quelle mesure il pouvait être utilisable pour le BUP. Il s'agit d'un thesaurus généraliste, destiné à des élèves de collège et lycée (plutôt niveau secondaire), toutes disciplines confondues. Cet aspect «généraliste» peut être considéré comme un avantage. Toutefois, dans notre cas, le public qui nous concerne est spécifiquement un public d'enseignants, d'étudiants scientifiques de bon niveau (CAPES, AGREG, DEA, ...) et, depuis deux ans, d'élèves de classes préparatoires dans le cadre des Travaux Personnels d'Initiative Encadrés (TIPE). Il s'agit donc dans tous les cas de «spécialistes» dans deux domaines : scientifique et éducatif, contrairement aux documentalistes des CDI qui ne sont que rarement scientifiques.

Dans son état actuel, le thesaurus MOTBIS :

- serait totalement à revoir du point de vue de *l'architecture logique* de la partie qui nous concerne (physique, chimie, technique, électronique, ...),
- propose un *niveau de spécificité* des mots retenus très variable d'un domaine à l'autre, mais globalement *très insuffisant* pour le public spécialisé défini ci-dessus.

Il aurait donc été indispensable, si nous avions voulu utiliser MOTBIS, d'élaborer de toutes pièces (ou presque) un ou plusieurs «microthesaurus» qui complèteraient MOTBIS dans notre domaine et compte tenu de la spécificité du public. C'est un travail de longue haleine, et qui ne pouvait être envisagé en dehors du groupe qui maintient ce thesaurus. Un contact pris avec le CNDP nous a appris qu'une nouvelle version du thesaurus était en préparation (début de l'année 1997) mais ne comprenait que des modifications restreintes.

L'analyse ci-dessus est valable pour de nombreux thesaurus qui ne correspondent pas à notre «cahier de charges» : ils sont soit trop généralistes, soit trop spécifiques (thesaurus spécialisés de bibliothèques techniques).

Les recherches «plein texte»

Les évolutions rapides dans le domaine technologique permettent maintenant d'envisager des solutions qui, il y a encore quelques années, auraient été impossibles. En particulier l'augmentation des capacités mémoire et de la rapidité des ordinateurs ont permis l'apparition de systèmes de gestion de bases de données permettant de rechercher n'importe quel mot dans un texte complet ou dans tous les champs de description d'une base.

Il est évident que de telles techniques ne sont exploitables que dans l'optique de bases de données informatisées.

Nous avons donc décidé de ne pas travailler à partir d'un thesaurus existant dans la mesure où nous n'en avons pas trouvé de conformes à nos objectifs et à notre public.

Nous n'avons pas non plus envisagé d'en élaborer un «sur mesure» à la fois parce qu'il s'agit d'un travail gigantesque qui aurait nécessité des moyens (financiers et disponibilité) dont nous ne disposons pas, mais aussi parce qu'il nous a semblé que les technologies actuelles rendaient cette option moins incontournable qu'il y a quelques années.

Les outils informatiques

Le choix se posait alors entre l'utilisation d'un logiciel de gestion de base de données commercial ou un logiciel adapté à nos besoins et écrit par nos soins. L'avantage de la première solution réside dans l'universalité de certains produits. Certains des collègues du groupe avaient d'ailleurs ébauché des bases de données avec divers produits (WORKS, FILMAKER, ENDNOTEPLUS, ...). Toutefois certains de ces logiciels coûtent relativement cher et aucun d'entre eux n'aurait probablement fait l'unanimité auprès des lecteurs du BUP. Il n'est pas possible de diffuser les logiciels dont les utilisateurs doivent alors disposer. Par ailleurs les problèmes de maintenance et d'évolution des versions nous ont paru redoutables.

La deuxième solution implique évidemment de disposer des compétences nécessaires. L'avantage est de pouvoir diffuser, en même temps que la base de données, le logiciel d'utilisation (sous Windows et sous MacOS), d'en rester maître, de pouvoir le faire évoluer en fonction de nos besoins, de développer parallèlement l'interface sur serveur *via* Internet.

LES GRANDES OPTIONS DE NOTRE TRAVAIL

Compte tenu des diverses réflexions exposées ci-dessus (utilisateurs potentiels, enseignants et étudiants spécialisés en sciences physiques, nécessité de construire à la fois une table quinquennale «papier» et une base de données informatisée) nous nous sommes orientés vers une solution «mixte» caractérisée par :

Sur le plan documentaire

- **Un point de vue documentaire** : même sans utiliser de thesaurus, le point de vue adopté est un point de vue *documentaire* et non *classificatoire*. C'est l'une des différences majeures par rapport à la table précédente. En particulier, nous n'avons pas utilisé comme mots clés les mots décrivant le «domaine» auquel appartient l'article : par exemple les articles de chimie organique ne sont pas indexés sous le mot «chimie organique», ils sont indexés à un niveau beaucoup plus fin selon le contenu de l'article (par exemple stéréoisomérisation ou aldéhyde, ...). On ne retrouvera donc pas, dans la table «papier» l'ensemble des articles de tel ou tel domaine regroupé derrière le terme correspondant. Un article ne serait indexé sous le mot clé «chimie organique» que s'il s'agissait d'une réflexion **SUR** la chimie organique ou d'un article d'histoire de la chimie organique par exemple.

Nous avons toutefois ajouté à la description des articles quelques données complémentaires : le «domaine», le point de vue et le niveau de l'article. Dans la table «papier» ces données servent à préciser le contenu de l'article mais ne sont pas des

«entrées» dans la liste alphabétique. Dans la base de données informatisée, on pourra par contre sélectionner des articles selon ces critères. On trouvera dans l'article suivant une description plus complète de ces données complémentaires.

- L'indexation se fait donc par mots-clés : contrairement à un thesaurus, cette liste de mots clés n'a pas été construite «a priori» mais de façon empirique au fur et à mesure de l'indexation. Les mots clés ont été choisis en se posant la question systématique : avec quel(s) mot(s) souhaiterais-je retrouver cet article ? Il ne s'agit pas nécessairement de mots du titre dans la mesure où le titre n'est pas toujours le meilleur indicateur du contenu.

Nous avons conservé certaines des idées mises en œuvre dans les thesaurus, en particulier quelques liens entre les mots clés.

Dans la mesure où l'indexation a été faite à plusieurs, nous avons dû, dans un deuxième temps faire un très important travail d'homogénéisation et de cohérence sur les mots clés. On trouvera dans l'article suivant («du bon usage de la table papier») quelques précisions sur ce choix des mots clés.

- Mots clés «papier» / mots clés «informatiques» :

- les mots clés destinés à la version «papier» sont au nombre de deux à trois pour des raisons techniques évidentes, chaque article apparaissant physiquement autant de fois qu'il y a de mots retenus pour son indexation ;

- les mots clés destinés à la recherche automatique sont au nombre de huit maximum (y compris les mots clés papier) ; on aurait pu ne pas limiter ce nombre ; toutefois il nous a semblé préférable de le faire pour indiquer clairement les principaux thèmes de l'article. Partant du principe que «trop d'informations tue l'information» nous avons pensé qu'il valait mieux huit mots clés bien choisis qu'une pléiade de mots n'ayant fait l'objet que de peu de réflexion.

Ces mots clés informatique apparaîtront également dans la table «papier» et remplaceront (sans doute de façon moins satisfaisante) les quelques lignes de commentaire des tables précédentes.

- Recherche dans la base de données informatique : la recherche se fera à partir de la liste finie de mots clés utilisés pour l'indexation et dans un champ spécifique réservé aux mots clés. Une deuxième option de recherche «en texte libre» permettra de chercher n'importe quel mot dans n'importe quel champ de description de l'article en particulier sur le titre ou dans un champ «résumé» qui ne sera encore que très imparfaitement rempli dans la première version de la base mais que nous espérons compléter au fur et à mesure des mises à jour.

Sur le plan informatique

Nous diffuserons la base de données informatisée à partir de la rentrée de septembre 1997 accompagnée d'un logiciel écrit par nos soins qui fonctionnera à la fois sous Windows et sous MacOS et fera l'objet de mises à jours régulières.

Cette base contiendra, outre l'ensemble des articles de la version «papier» y compris l'année 1996, tous les articles des numéros spéciaux informatiques (journées IPSP et photocopiés spéciaux), les sujets de CAPES et d'AGREG indexés individuellement ainsi que les sujets des baccalauréats technologiques.

La version de la base consultable sur le serveur aura une ergonomie voisine.

C'est Olivier KEMPF et Claude LACOMBE qui ont pris en charge l'énorme travail de conception et de réalisation de la base de données et des outils de consultation. Outre leurs compétences tant dans le domaine de l'informatique que de la physique et de la chimie, la qualité de leurs propositions, leur rapidité d'exécution, toute l'équipe a apprécié de pouvoir discuter avec des «physiciens-informaticiens» à l'écoute des problèmes qui leur étaient posés et ne répondant pas systématiquement que c'est impossible... à cause de l'informatique !

Rappelons enfin que cette table, malgré tout le soins que nous y avons apporté, comportera sans doute un certain nombre d'erreurs. D'avance nous nous en excusons, mais surtout nous vous demandons instamment de nous les signaler afin que la version «disquette» de la base de données puisse bénéficier de ces échanges : ce n'est pas l'un des moindres avantages d'une version informatique que de pouvoir assurer une mise à jour régulière : nous comptons sur vous et d'avance nous vous en remercions.

ET MAINTENANT ?

Notre travail n'est pas terminé ! Après un an d'échanges, de discussions, de travaux en commun, le groupe est bien rodé et envisage de continuer à réfléchir à de nouvelles perspectives : indexation de l'ensemble de la collection depuis sa création, transfert des archives sur supports informatiques, élaboration d'outils appropriés... Il serait également intéressant de pouvoir disposer, avec les mêmes outils, des tables des matières ou des index d'autres revues scientifiques à notre disposition soit dans les CDI des établissements, soit dans notre documentation personnelle. Enfin l'indexation du BUP dans les grandes banques de données nationales et internationales sera également à étudier. Les idées ne manquent pas, le temps souvent !...

Sommaire

Éditorial : de la table quinquennale à la base de données

Du bon usage de la table «papier» et de ses index

Index des articles (classés par mots clés)

Index des auteurs (alphabétique)

Connaissez-vous le BUP ?

Autres index

Cette table quinquennale a été élaborée par le groupe «Base de données» de l'Union des Physiciens

Conception des bases de données, gestion informatique

Olivier KEMPF - IUFM d'Alsace - Strasbourg

Élaboration des outils informatiques de consultation

Claude LACOMBE - Lycée Champollion - Grenoble

Indexation, outils documentaires

Josette CARRETTO - *Préparation CAPES-chimie* - Université Paris XI - Orsay

Odile CARRIER - Lycée Paul Valéry - Paris

Paul GOUVERNEUR - Collège Saint-Pierre - Bruxelles

Olivier KEMPF - IUFM d'Alsace - Strasbourg

Claude LACOMBE - Lycée Champollion - Grenoble

Bernard MONTFORT - Département de Chimie - IUT de Besançon

Gérard MORA - IUFM de Bretagne - Lycée Mendès-France - Rennes

Alain NOEL - Rectorat - Rouen

Jean-François PIGNAL - Lycée Brequigny - Rennes

Marcelle PROFIT - Retraitée - Clermont-Ferrand

Monique SCHWOB - Lycée Georges de la Tour - Metz

François VILLEDARY - Lycée Montaigne - Paris

Coordination

Monique SCHWOB - Rédacteur en chef