

APPORT DE LA MICRO INFORMATIQUE EN BIOLOGIE : LE PROGRAMME XPER ET L'IDENTIFICATION ASSISTEE PAR L'ORDINATEUR (I.A.O.)

par J. LEBBE et R. VIGNES

Introduction

Un peu d'histoire

Les progrès récents de la micro-informatique apportent de nouveaux outils qui améliorent grandement la communication, l'éducation et l'exploitation des connaissances.

Les premiers micro-ordinateurs ont vu le jour vers la fin des années 70. Avec eux, la puissance informatique est devenue disponible pour tous, pour un faible investissement, et de nombreux programmes ont déjà révolutionnés certaines méthodes de travail. Parmi eux, les traitements de texte simplifient le travail de secrétariat en éliminant des tâches répétitives. Les tableurs, ou feuilles de calcul électronique, permettent de résoudre de nombreux calculs fastidieux. Enfin les programmes de gestion de fichiers simplifient et remplacent les fiches traditionnelles, classeurs d'adresses, fichiers clients, etc...

En 1985, la puissance offerte par les micro-ordinateurs permet la diffusion d'applications qui étaient jusqu'alors réservées aux gros systèmes informatiques. Ainsi de nouveaux types de programmes tels que les —SYSTEMES EXPERTS— ou les systèmes d'I.A.O. voient le jour sur les ordinateurs individuels.

L'utilisation de ces logiciels peut bouleverser les processus d'identification dans tous les domaines de la biologie : identification de taxons en systématique, de maladies en médecine ou en agromonie, de substances chimiques en Pharmacologie, etc...

Les systèmes-experts et d'I.A.O., leurs applications

Il existe de nombreux types de programmes regroupés sous le nom de —SYSTEMES EXPERTS—. Il s'agit de systèmes de gestion de bases de connaissance fournies à l'ordinateur sous forme de faits ou de règles. Tous ces systèmes se caractérisent par leur possibilité de déduire des faits ou des règles qui n'ont pas été introduits comme tels dans la base de connaissance.

Xper est l'un de ces logiciels. Il fonctionne sur les micro-ordinateurs les plus répandus : APPLE, COMMODE 64, IBM PC, APRICOT, OLIVETTI M24, etc...

Les applications de tels programmes sont très nombreuses, citons l'Enseignement Assisté par Ordinateur (E.A.O.) et l'Identification Assistée par Ordinateur (I.A.O.) qui s'adressent à toutes les opérations de diagnostic tel que le diagnostic médical ou l'identification de pannes ou de plantes.

Le Programme Xper

Xper est un logiciel permettant à la fois la création et l'utilisation de bases de connaissance. Sa facilité d'emploi le rend utilisable par tous, sans qu'aucune connaissance en informatique soit nécessaire. Les données sont fournies à l'ordinateur sous forme de textes clairs, sans aucun codage, et restent toujours accessibles pour être consultées, corrigées, complétées. Une fois la base de connaissances créée, Xper permet de l'étudier de multiples façons : recherches multi-critères, calculs de distances, extraction et comparaison de groupes. Mais surtout Xper fournit un moyen de déterminer incomparable et utilisable par tous c'est l'I.A.O..

Structure d'Xper

Xper s'adresse à tous les domaines car il offre à la fois le moyen de créer et d'utiliser une base de connaissance. Chaque spécialiste peut donc faire sa propre application.

Le programme comporte quatre modules indépendants :

- Un EDITEUR qui permet de créer, consulter, corriger une base de connaissance, ainsi que de faire des recherches multicritères, des comparaisons et des calculs de distances.
- Un DETERMINEUR. C'est le module d'I.A.O.
- Un RÉORGANISATEUR qui permet de modifier l'ordre de toutes les données, et de regrouper ensemble plusieurs bases.
- Un IMPRIMEUR qui affiche à l'écran ou sur papier le contenu de la base de connaissance selon différents choix.

Structure d'une base de connaissance

Une base de connaissance se présente grossièrement comme une matrice formée par une liste d'individus décrits par des variables qualitatives indépendantes pouvant avoir chacune jusqu'à 14 modalités. Toutes les données, individus, variables, modalités, sont entrés sous forme de texte et restent toujours accessibles en clair. Le codage effectué par Xper est réalisé automatiquement et invisible pour l'utilisateur.

Les individus ou lignes de la matrice représentent ce que l'on veut décrire et ultérieurement déterminer. Il s'agit par exemple des différents taxons en systématique.

Les variables ou colonnes de la matrice correspondent à ce que l'on veut décrire sur l'ensemble des individus. Xper utilise des variables qualitatives ayant chacune jusqu'à 14 modalités ou états possibles. Ces modalités sont indépendantes les unes des autres.

En fait la matrice des données peut être très complexe. Non seulement chaque variable peut avoir entre 2 et 14 modalités, mais la description de certaines variables peut dépendre de celles d'autres variables. Par exemple, dans le cas d'une base de connaissances sur les Insectes, il est inutile de répondre aux variables décrivant les ailes chez les individus qui n'en ont pas. On peut ainsi définir des variables mères et des variables filles qui peuvent à leur tour avoir des filles et ainsi de suite. La matrice n'est pas plane mais représente une superposition de matrices.

Les modalités indépendantes, le codage par classes des données quantitatives, et la filiation des variables entre elles permettent de formaliser des données même très complexes.

Un des points forts d'Xper est de fournir un mode de création, l'éditeur, permettant à toute personne sans aucune connaissance informatique de créer sa propre base de connaissance.

Une fois la base réalisée, elle peut être utilisée pour de multiples recherches, mais surtout dans bien des cas, elle remplace la nécessité de créer des clés de détermination.

La détermination s'effectue pas à pas en deux temps. Premièrement, l'utilisateur choisit parmi la liste des variables celle à laquelle il souhaite répondre. Puis, dans un deuxième temps, il indique la ou les modalités correspondant à la description de son spécimen.

Ces opérations se renouvellent jusqu'à l'obtention soit de l'identification, soit d'une discrimination impossible car les éléments de réponses ne suffisent pas à définir un seul individu, soit d'une combinaison de caractères imprévue dans la base.

Les avantages de l'identification assistée par ordinateur, offerte par Xper, sont nombreux :

— L'utilisateur a le choix de sa stratégie de détermination. Il peut choisir les caractères qu'il veut décrire, et donc n'est jamais arrêté par une donnée manquante ou un caractère difficile à observer sur son spécimen.

L'identification assistée par ordinateur

La base de connaissances étant saisie, Xper offre immédiatement le moyen d'effectuer des déterminations avec de grandes facilités par rapport à une clé classique. Le travail supplémentaire, et souvent long de la construction d'une clé, est inutile, et la mise à jour des connaissances est immédiate.

— Le doute est pris en compte. Cette possibilité permet d'utiliser des caractères plus flous que dans une clé, mais pouvant dans bien des cas être utiles.

— Le retour en arrière dans la détermination est très facile et ne perd pas les réponses déjà effectuées.

— Xper peut éliminer automatiquement les variables et les modalités non discriminantes, ou rechercher les variables les plus discriminantes pour identifier un individu particulier.

— La démarche de la détermination est mémorisée.

Xper justifie son comportement tout au long de la détermination, et permet d'apprécier la validité des résultats obtenus. A chaque étape de la détermination, l'utilisateur a accès à la liste des individus restant compatibles avec la description du spécimen, et aux raisons de l'élimination des autres individus. Il peut aussi calculer le nombre de différences entre la description du spécimen à identifier et chaque individu de la base.

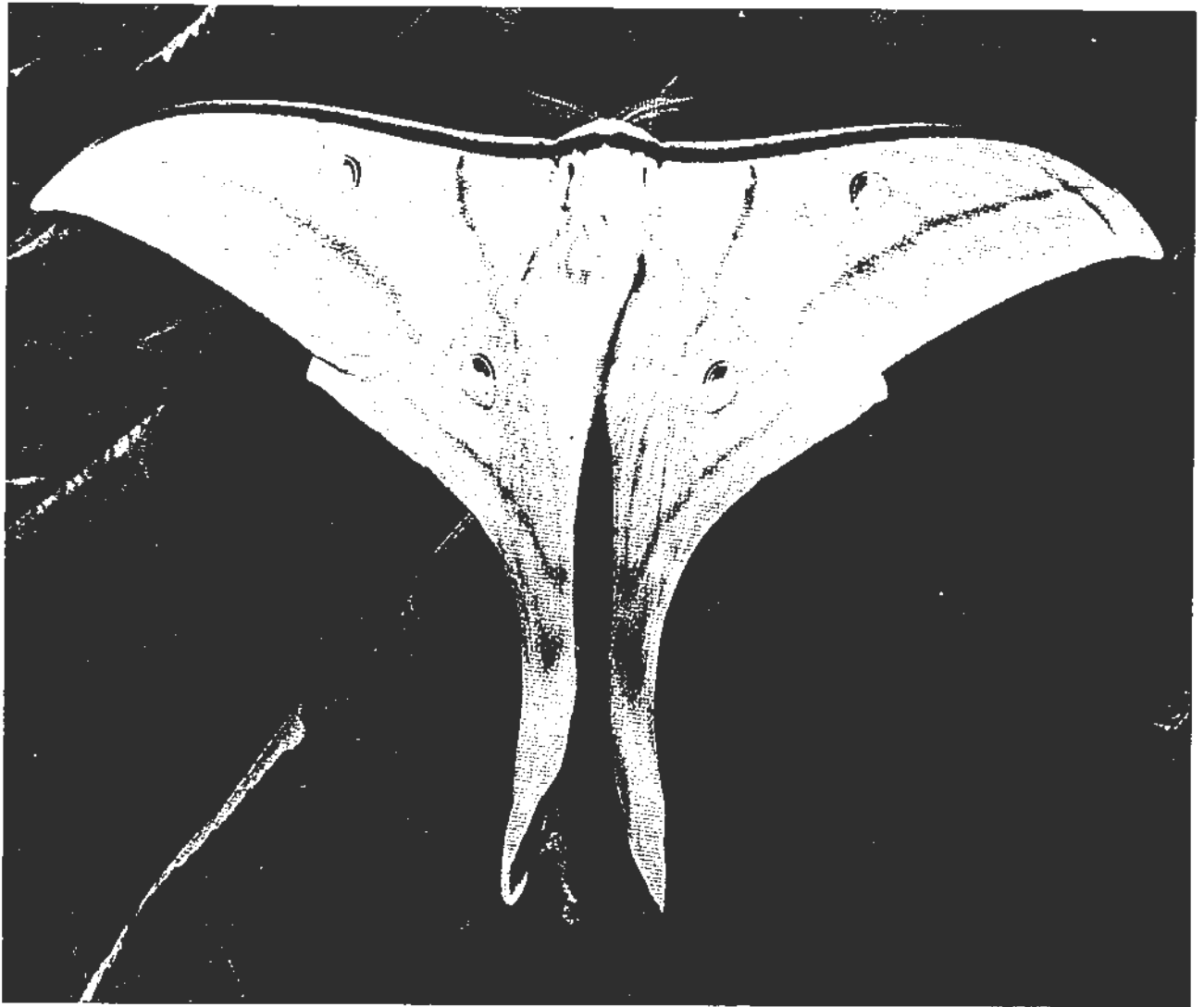
Xper apporte donc à tous ceux qui ont besoin d'effectuer des diagnostics ou des déterminations un outil très facile d'emploi et de remise à jour immédiate.

Conclusion

Xper a été présenté dans plusieurs laboratoires du Muséum d'Histoire Naturelle, à l'Institut Pasteur, à l'I.N.R.A., et dans des stations biologiques où il a suscité un grand intérêt. Plusieurs applications ont été réalisées ou sont en cours de réalisation dans différents domaines. Lors d'un D.E.R. de botanique tropicale Xper a montré la grande facilité de détermination que procure un tel programme pour les arbres de la Guyane Française. La création d'une base de connaissance sur les champignons de France est actuellement en cours.

L'utilisation de la micro-informatique et de programmes tels qu'Xper permet le stockage en clair de toutes sortes de données et facilite leur diffusion ainsi que leur étude approfondie. La création de bases de connaissance par les spécialistes offrent ensuite à ceux ayant la nécessité de déterminer un moyen facile à utiliser et à remettre à jour.

N.D.L.R. : Nous pensons que cet article suggèrera à certains d'entre nous de concevoir un programme facilitant l'identification rapide des insectes dans des travaux d'inventaire de faunes et de dépouillement de résultats de captures au moment du tri des espèces.



Mâle d'*Actias selene*. Cliché G. BOULOUX