ISSN 0206-0477. **БАЗЫ ДАННЫХ И КОМПЬЮТЕРНАЯ ГРАФИКА В ЗООЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ**
СПб., 1997 (ТРУДЫ ЗООЛОГИЧЕСКОГО ИНСТИТУТА РАН, Т. 269)

RUSSIAN ACADEMY OF SCIENCES
PROCEEDINGS OF THE ZOOLOGICAL INSTITUTE
St. PETERSBURG, 1997, VOL. 269

УДК: 59.08:681

A.F. Alimov, A.L. Lobanov, O.N. Pugachev

Zoological Institute, RAS, St.Petersburg, Russia

COMPUTER TECHNOLOGY IN THE ZOOLOGICAL INSTITUTE RUSSIAN ACADEMY OF SCIENCES

The use of computer technology in the Zoological Institute coincided with the PC appearance in the former Soviet Union. Large arrays of information or data bases in the previous computer generation were nearly absent, except data on bird banding (Kaliningrad Biological Station). Since the PC appearance was 3-4 years late as compared with western states, we then had to start with different relatively modern software, which had been in use for a long time. As a result problems that were typical of each organization using computer technology appeared immediately. The other problem was that zoological information was usually structured, but not formalized. Friendliness of the modern software as compared with that of the previous computer generation leads to easiness in use for different users even for nonspecialists in information technology and finally to information "erosion". Software diversity determines the appearance of incompatible databases as a result of different information structure and different approaches to database creation.

Diverse information about enormous number of animal species is accumulated in the Zoological Institute which are stored in collections as well as in the form of knowledge of zoologists specialized in different systematic groups. The problems which were mentioned above defined the necessity to have a special computerization program of the zoological investigations in the Zoological Institute. This program consists of several stages. The first stage includes database development which includes data structure and specific software development, keyboard input. The second one is the development of the information retrieval system. The third one

includes development of the expert systems for the estimation of ecosystem state and anthropogenic impact. It was obvious that keyboard input will be the most prolonged stage requiring considerable investments.

Zoological database development is based on the following main principles:

1. The base information unit is a species.
2. The system of hierarchy sets of information units (formalized representation of the existing animal classifications) has to be the structural base of a zoological database.
3. The application program has to support flexible possibilities for changes in this system of hierarchy sets.

The system of hierarchy classification was used for the first time based on the relational approach by A.L. Lobanov. The special software and original list representation were produced. The readers can become acquainted with the results of this approach in several publications of this issue.

It became clear during collection database development that large array of information and increasing zoological database structure complexity in the line "collection database - faunistic database - ecological database" can exceed the facilities of the widely spread relational database management systems. Our experience in using different data base management systems leads us to the conclusion that the postrelational network database management is more suitable for the future database development because of the complexity of information structure. Advantages of the postrelational network model become obvious when we have a large array of information and its complex hierarchical structure. As this model has no file structure, common fields and coding procedure the performance of such kind of information system increases (for more detailed comparative analysis of network and relational approach see e.g. Abbot et al.,1985; Alimov et al.,1993).

The storage and processing of a primary parasitological data (localities, host with special features, parasite species with localization, number, automatic calculation of prevalence, intensity, abundance etc.) was realized with postrelational network database management.

Data model of the program "Notebook for parasitology" is given in Fig. 1.

Record name is COLL - list of collectors, item name is TEXT - collector name.

Record name is FISH - list of studied fish species, item name is TEXT - fish name, item name is GENUS - fish genus, item name is SPECIES - fish species name.

Record name is CONCR - features of studied fish specimens, item name is NUMB - fish number according to field notebook, item name is DAT - data of investigation of fish specimen with format mm/dd/yy, item name is LONGAB - fish length AB, item name is LONGAC - fish length AC , item name is LONGAD - fish length AD, item name is WEIGTH - fish weight, item name is AGE - fish age, item name is SEX - fish sex.

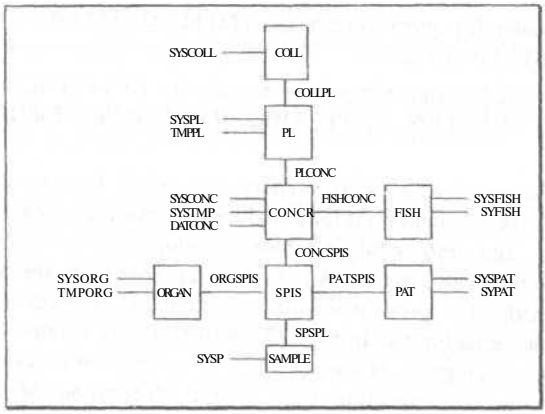

Fig. 1 Model of postrelation network database for "Notebook for Parasitology"
Модель управления сетью базы данных "Паразитологкческий справочник"

Record name is SPIS - number of one parasite species on one host specimen, item name is NUMB - parasite number.

Record name is PAT - list of parasite species, item name is TEXT - parasite genus, item name is SPECIES - parasite species name, item name is AUTSH - author, year.

Record name is PL - list of localities where parasitological investigations were done, item name is TEXT - locality name, item name is WIDTHD - latitude (degree), item name is WIDTHM - latitude (minute), item name is WIDTHS - latitude (second), item name is WNAM - latitude name (north/south), item name is LONGD - longitude (degree), item name is LONGM - longitude (minute), item name is LONGS - longitude (second), item name is LNAM - longitude name (east/west),

item name is CHK - field which use for pooling of different localities in one query.

Record name is SAMPLE - list of slides with parasites which were collected from one host specimens, item name is NO - slide number, item name is COMMENT - comment text to slide.

Record name is ORGAN - list of host organs, item name is - TEXT - organ name.

There are three major groups of sets:

1. Sets responsible for access to a database record directly - SYSCOLL, SYSPL, SYSPAT, SYSORG, SYSCONC, SYSTMP, SYSFISH, SYSP (type 1:n);
2. Sets responsible for query formation - TMPPL, DATCONC, SYPAT, TMPORG, SYFISH, (type l:n);
3. Sets responsible for relationships between records - COLLPL (type m:n), PLCONC, CONCSPIS, SPSPL, ORGSPIS, PATSPIS, FISHCONC (type 1: n).

This program allows to store primary data which have never been published but are very important for a study, e.g. parasitic communities.

Faunistic and ecological database development is linked to geographical retrieval systems. An original approach to the use of coordinates without a special manipulation within geographical retrieval system was elaborated in our Institute. Thus further data manipulation is based on the zoological database, which is a real time saver. The difficulties with PC upgrading lead to the elaboration of special application programs, which allow to use geographical retrieval systems with the minimum requirements to hardware. The other specific feature is that the algorithm for data manipulation has been developed together with zoologists. It provides a means for data representation in the form usual for zoologists. Several papers in this issue deal with this approach.

Next field for use of computer technology is computer identification systems, which provides a means for intensification of routine zoological procedures. The application programs are based on data from usual "dbf" files. There are original algorithms for the estimation of feature diagnostic value and a relatively simple interface. The main problem is feature formalization, which is analogous to multientry identification key development. This kind of identification keys has not yet gained wide acceptance at least in our country. Progress in this area of study depends on enthusiasm of taxonomists. The readers can become acquainted with the results in this field in several papers of this issue.

The material identification and primary data securing take a lot of time, which is connected with viewing a huge number of objects, measurings, and comparison. Thus image analysis systems seem to be very important in this field. Special efforts were made during last year. As a result we have this kind of equipment in three departments. This field of use of computer technology in zoology seems to be highly promising taking into account the possibilities to include images into the databases.

Computer technology is in use in the main fields of zoological investigations in our Institute now. The time is right to co-ordinate activity in this field more strictly than it was done before.

А.Ф. Алимов, А.Л. Лобанов, О.Н. Пугачев

Зоологический институт РАН, Санкт-Петербург, Россия

ИСПОЛЬЗОВАНИЕ КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ В ЗООЛОГИЧЕСКОМ ИНСТИТУТЕ РАН

Описывается программа компьютеризации зоологических исследований, осуществляемая в Зоологическом институте РАН. Первый этап - разработка баз данных (БД), их структуры и выбор адекватного программного обеспечения, ввод данных в БД. Второй этап - развитие информационно-поисковых систем, а третий - экспертных систем для оценки различных параметров биоразнообразия и состояния экосистем. Каждое из этих направлений в последние годы получило развитие в институте. В рамках первого и второго направлений наиболее широко используется иерархическая классификация на основе реляционных СУБД (систем управления базами данных). Кроме того, обосновано использование постреляционного сетевого программного обеспечения по управлению БД на примере "Дневника паразитолога", включающего хранение и обработку первичных данных. Третье направление представлено целым комплексом программных средств на основе реляционных СУБД для автоматизации работы зоолога-систематика. Создаваемый комплекс позволяет оценивать диагностическую ценность различных признаков и разрабатывать многофункциональные диагностические ключи.