

ИТОГИ РАЗВИТИЯ ПРОЕКТА ЗООИНТ И ЕГО ДАЛЬНЕЙШИЕ ПЕРСПЕКТИВЫ

И.С.Смирнов, А.Л.Лобанов, А.Ф.Алимов, С.Г.Медведев, А.А.Голиков
Зоологический институт РАН,
199034, Россия, С.-Петербург, Университетская наб., д.1
smiris@zin.ru

THE RESULTS OF DEVELOPMENT OF THE PROJECT ZOOINT AND ITS FUTURE PERSPECTIVES

I.S.Smirnov, A.L.Lobanov, A.F.Alimov, S.G.Medvedev, A.A.Golikov
Zoological Institute of RAS, University nab., 1, St.-Petersburg, Russia, 199034
smiris@zin.ru

The work on a computerization of main processes of accumulation and analysis of the collection, expert and literary data on a systematics and faunistics of various taxa of animal (a basis of studying of a biological diversity) was started in the Zoological Institute in 1987. In 1991 the idea of creating of the software package, ZOOlogical INTegrated system (ZOOINT) was born. ZOOINT could provide a loading operation about collections and simultaneously would allow to analyze the accumulated data with the help of various queries.

During execution, the project ZOOINT was transformed slightly and has given results a little bit distinguished from planned earlier, but even more valuable. In the Internet the site about the information retrieval system (IRS) ZOOINT was built also (http://www.zin.ru/projects/zoint_r/index.html). The implementation of remote access to the taxonomic information, with possibility to work with databases (DB) of the IRS ZOOINT in the on-line mode was scheduled. It has required not only innovation of computer park of the developers and users, but also mastering of new software: language HTML, operating system of Windows NT, and technology of Active Server Pages (ASP).

One of the serious problems of creating of databases and the IRS on zoology is the problem of representation of hierarchical classification. Building the classifiers, specialized standard taxonomic databases, which have obtained the name ZOOCOD solved this problem. The lately magnified number of attempts of creating of taxonomic electronic lists, tables and DB has required development of some primary rules of unification of zoological systematic databases. These rules assume their application in institutes of the biological profile, in which the processes of a computerization are very slowly, and the building of databases is in the most rudimentary state. These some positions and the standards of construction of biological (taxonomic) databases should facilitate dia-

logue of the biologists, application in the near future of most advanced technologies of development of the DB (for example, usage of the XML platform) and, eventually, building of the modern information systems.

The work on the project is carried out at support of the RFBR grant N 02-07-90217; programs "The Information system on a biodiversity of Russia" and Project N 15 "Antarctic Regions".

Работа по компьютеризации основных процессов накопления и анализа коллекционных, экспертных и литературных данных по систематике и фаунистике различных таксонов животных - основы изучения биологического разнообразия - была начата в Зоологическом институте с появлением в 1987 г. терминала БЭСМ-6, а в 1988 г. - СМ-1420. Приобретение персональных компьютеров в 1989-90 гг. ускорило разработку прикладных программ для зоологов [1, 6, 8, 10, 15].

В 1991 г. родилась идея создания пакета программ, Зоологической Интегрированной системы (ЗООИНТ), которая могла бы обеспечить ввод информации о коллекциях и одновременно позволила бы анализировать накопленные данные с помощью различных запросов [9].

В 1993 г. эта идея информационно-поисковой системы (ИПС) по зоологии получила поддержку Российского Фонда Фундаментальных Исследований (проект N 93-04-21216 "Система компьютерной интегрированной обработки данных по биоразнообразию животных (ЗООИНТ)") [3, 10].

В 1996 г. разработка проекта была продолжена при поддержке РФФИ (грант № 96-07-89086). Проект получил название "Развитие интегрированной системы ZOOINT для использования зоологических баз данных" [4, 11, 12, 14]. В силу сложившихся в зоологических учреждениях России предпочтений на первом этапе создания ИПС нами была выбрана система управления базами данных (СУБД) dBase III+, которая уступила место FoxBase, и, наконец, последняя трансформировалась в FoxPro, сначала под MS DOS, затем под Windows. К сожалению, система FoxPro, приобретенная корпорацией Microsoft, практически завершила свое развитие. Тем не менее, освоение этих относительно несложных СУБД позволило зоологам, не привлекая сторонних специалистов и полагаясь только на свои силы, приступить к проектированию и созданию зоологических БД. При низком уровне заработной платы, который существует в академических учреждениях зоологического профиля, привлечь со стороны высококвалифицированных программистов практически было невозможно.

В ходе выполнения, проект ЗООИНТ немного трансформировался и дал результаты, несколько отличающиеся от планировавшихся ранее, но даже более ценные. После появления первой пилотной версии ИПС

возникла настоятельная потребность в связывании в единую сеть, поначалу компьютеров разработчиков, а затем и институтских подразделений, которые могли бы пополнять БД и использовать уже накопленные данные в своих целях. Это был первый толчок к дальнейшему развитию системы ЗООИНТ. Вторым мощным толчком к развитию компьютерных зоологических систем стало появление мультимедийной технологии, которая резко расширяет возможности накопления, хранения и использования зоологической информации (особенно изображений – фото и видео, сейчас уже и трехмерных).

К 1999 г. система ЗООИНТ поддерживалась СУБД FoxPro для DOS в локальной сети из 7 компьютеров. Был также создан сайт об ИПС ЗООИНТ в сети Интернет (http://www.zin.ru/projects/zooint_r/index.html).

В 1999 г. по итогам конкурса РФФИ был поддержан проект "Информационно-поисковая система ЗООИНТ по биоразнообразию животных" № 99-07-90315 [2, 7, 15]. Основной целью проекта служило продолжение разработки комплекса программных средств для работы с классификацией животного мира, которые стали базой для создания конкретных БД и ИПС систематического, экологического и коллекционного характера.

В последнем проекте 1999 г. усилия были направлены на интенсивное развитие локальной общеинститутской сети, которая позволила бы объединить усилия специалистов-зоологов по вводу данных по разным группам животных и, главное, воспользоваться уже накопленной в базах данных информацией. Благодаря усилиям администрации, Институт, наконец, получил постоянный и достаточно быстрый доступ в Интернет. Была запланирована реализация удаленного доступа к таксономической информации, с возможностью работать с базами данных ИПС ЗООИНТ в режиме on-line. Это потребовало не только обновления компьютерного парка разработчиков и пользователей, но и освоения новых программных средств: языка HTML, операционной системы Windows NT, технологии Active Server Pages (ASP) и др. [2].

По мере разработки проекта отдельные аспекты ЗООИНТ превратились в самостоятельные направления. Многолетняя работа над коллекционными базами данных дала целое дерево ответвлений. Центральный ствол представлен интегрированной системой ЗООИНТ. Основное внимание при развитии ИПС ЗООИНТ было уделено коллекционным данным. Одна из ветвей ИПС ЗООИНТ дала систему ОКЕАН, которая коллективно используется и совершенствуется в лабораториях морских исследований и ихтиологии. ИПС ОКЕАН ориентирована на ввод и хранение данных по коллекциям морских животных [13]. Мощное ответвление представлено системой по мировой фауне блох PARHOST, развиваемой С.Г.Медведевым с коллегами. Другая ветвь системы ЗООИНТ развивается с участием А.Ф.Алимова, В.Е.Панова

и их коллег. Это – информационно-поисковая система INVADER, которая служит для накопления и анализа данных об организмах-вторженцах: <http://www.zin.ru/projects/invasions/rus/index.html#> [7]. Параллельно с сайтом ИПС ЗООИНТ создавались сайты по биоразнообразию, отражающие широкий спектр информации о двух важных группах насекомых – жуках и блохах, это соответственно – www.zin.ru/animalia/coleoptera и – www.zin.ru/animalia/siphonaptera.

Продолжающиеся развиваться сайты по жукам и блохам, по насыщенности информации и углубленной проработке темы не имеют аналогов в русском Интернете, о чем говорит высокая посещаемость данных страниц (в 2001 г. на сайте жуков отмечено около 77 тысяч посещений, блох – около 76 тысяч). Успешное развитие ЛВС Зоологического института, подключение к сети Интернет и работа над проектом ЗООИНТ и ее результаты инициировали разработку новых зоологических информационных проектов по различным группам животных и некоторые из них получили в 2000-2001 гг. гранты РФФИ: 00-07-90304, 00-07-90287, 00-07-90288, 01-07-90284, 02-07-90105, соответственно, это – пресноводные рыбы, простейшие, пресмыкающиеся и насекомые. Пресноводные рыбы России – проект, которым руководит Н.Г.Богущая (<http://www.zin.ru/Animalia/Pisces/index.html>); Протисты – электронная коллекция, проект, разрабатываемый А.О.Фроловым с коллегами (<http://www.zin.ru/Animalia/Protista/index.html>); Таксономическая и коллекционная интерактивная база данных по мировой фауне насекомых ZInsecta (руководитель информационного проекта В.А.Кривохатский – <http://www.zin.ru/projects/ZInsecta/index.html>); Веб-сайт по жукам (рук. А.Г.Кирейчук – <http://www.zin.ru/animalia/coleoptera>).

Развивает самостоятельно сайт по круглым червям А.Ю.Рысс (Класс Nematoda – <http://www.zin.ru/Animalia/Nematoda/index.html>).

Представляет интерес проект ЭКОАНТ – информационная система по ЭКОлогии морского бентоса и прибрежным птицам АНТарктики (<http://www.zin.ru/projects/ecoant/index.html>), где в качестве примера выставлены коллекционная база данных по антарктическим беспозвоночным – змеехвосткам и данные по прибрежным птицам.

В 2001 г. Зоологический институт выиграл по конкурсу государственную поддержку по программе "Информационная система по биоразнообразию России", которая успешно развивается.

Одной из серьезных проблем создания баз данных и ИПС по зоологии является проблема представления иерархической классификации. Эта проблема была решена путем создания классификатора, специализированной таксономической базы данных, получившей название ЗООКОД или ZOOCOD. Этот стандарт был разработан в конце 80-х годов в ЗИН РАН А.Л.Лобановым для преобразования зоологических иерархических классификаций в плоскую реляционную таблицу, таким

образом, чтобы все сведения, которые содержатся в систематическом списке, были отражены в соответствующих полях классификатора. Усовершенствование и многократные демонстрации и испытания стандарта ZOOCOD привели к его широкому распространению (Зоологический и Ботанический институты, Институт проблем эволюции и экологии, Московский государственный педагогический институт, Нижегородский государственный университет и др.) и появлению очередной версии - ZOOCOD3. Практически во всех создаваемых в ЗИН РАН базах данных есть таксономические блоки. В большинстве этих блоков для отражения иерархических классификаций в плоских реляционных таблицах используется концепция ZOOCOD2, подробно описанная ранее [5].

Увеличившееся за последнее время число попыток создания таксономических электронных списков, таблиц и БД потребовало разработки некоторых первичных правил унификации систематических баз данных. Эти правила предполагают их применение в учреждениях биологического профиля, в которых очень медленно идут процессы компьютеризации, и создание баз данных находится в самом зачаточном состоянии.

Эти некоторые положения и стандарты построения биологических (таксономических) баз данных должны облегчить общение биологов, применение в ближайшем будущем наиболее продвинутых технологий развития БД (например, использование платформы XML) и, в конечном счете, создание современных информационных систем:

1. Должна использоваться реляционная модель баз данных.
2. В соответствии с требованиями нормализации баз данных информационные системы должны включать отдельные таблицы по каждому аспекту сведений о животных или растениях.
3. Каждый элемент сведений должен храниться в конкретной системе только в одной записи одной таблицы. Это правило влечет за собой широкое использование словарей и классификаторов (или тезаурусов - словарей с иерархической структурой и родо-видовыми отношениями между понятиями).
4. В основе системы должны лежать классификаторы животных, растений, грибов, бактерий и других царств живых организмов.
5. Часто используемые биологами данные о географическом распространении организмов должны быть строго разделены на отдельные аспекты (административно-территориальный, физико-географический, биогеографический, ландшафтный, природно-зональный и т.п.). Термины каждого аспекта должны быть зафиксированы в классификаторах.
6. Добиться ведения разных баз данных в одной СУБД невозможно даже внутри одного института и к этому не нужно стремиться. Важно только использовать профессиональные коммерческие СУБД, имеющие

проверенные механизмы импорта и экспорта. В качестве критерия пригодности СУБД предлагается использовать формат DBF (dBaseIII).

7. Для обеспечения простоты обмена данными между СУБД, работающими в разных версиях операционных систем, следует ограничить длину имен файлов 8 символами, длин расширений имен 3 символами, а названия полей баз данных 10 символами; при этом использовать во всех этих именах только латинские буквы, цифры и символ подчеркивания (при этом не стоит использовать разницу между заглавными и строчными буквами, если какая-то система такую возможность поддерживает). В системах, где длина символьного поля не ограничивается или может быть очень большой, не делать длину поля более 254 символов.

8. Если для представления иерархии не используется стандарт ZOOCOD, то все же нужно стремиться к соблюдению некоторых правил, облегчающих переход к этому стандарту:

а) для указания ранга таксона или понятия использовать только одно поле таблицы, приводя в нем коды или полные названия рангов;

б) иметь поле с уникальным кодом таксона и использовать эти коды для связи с таблицами фактов о таксонах;

в) в отдельном поле отмечать факт синонимии, указывая уникальный код валидного названия для каждого синонимичного;

г) иметь поле для указания систематического порядка записей (упорядочение по нему должно создавать аналог "систематического списка", который привычен и необходим большинству биологов);

д) наличие двух полей, отражающих иерархию способом, принятым теперь в приложениях для Windows (поля для отображения иерархии методом TreeView), ни в коей мере не отменяет необходимость пунктов а) и г), так как эти два поля не отражают истинные ранги таксонов и нужный порядок среди дочерних таксонов одного родительского.

Авторы проекта считают, что данная разработка позволит не только облегчить труд зоологов, обеспечить экологов легко доступной таксономической информацией, но и поднимет биологические исследования на новый уровень. Трудно переоценить роль данной информационной системы для сбора и верификации колоссальной информации, которая сосредоточена в коллекциях Зоологического и других институтов и практически малодоступна для специалистов других областей науки и народного хозяйства.

Работа по проекту осуществляется при поддержке гранта РФФИ № 02-07-90217; программы "Информационная система по биоразнообразию России" и проекта № 15 "Антарктика".

Литература

1. Алимов А.Ф. Компьютеризация биологического учреждения (на примере Зоологического института РАН)// Информационно-поисковые системы в зоологии и ботанике, Труды Зоологического института РАН, т. 278, Санкт-Петербург, 1999: 6,7.
2. Алимов А.Ф., Смирнов И.С., Рысс А.Ю., Дианов М.Б., Лобанов А.Л., Голиков А.А. Современные биологические электронные публикации: коллекции, идентификационные системы и базы данных// Информационные и телекоммуникационные ресурсы в зоологии и ботанике. Тезисы 2-го международного симпозиума. СПб. 2001. с. 5-19.
3. Lobanov A., Sokolov E., Smirnov I. ZOOINT - an integrated system for zoological data bases// ADBIS'94. Proceedings of the International Workshop on Advances in Databases and Information Systems. May 23-26, 1994. Moscow, 1994. P. 270,271.
4. Лобанов А.Л., Смирнов И.С., Алимов А.Ф., Дианов М.Б., Степаньянц С.Д. Интегрированная информационная система ЗООИНТ: современное состояние и перспективы. Отчетная научная сессия по итогам 1996 года. Тезисы докладов. 8-10 апреля 1997. Изд. Зоологического института, 1997: 23-25.
5. Лобанов А.Л., Смирнов И.С. Принципы построения и использования классификаторов животных в стандарте ZOOCOD// Базы данных и компьютерная графика в зоологических исследованиях, Труды Зоологического института, т. 269, Санкт-Петербург, 1997: 66-75.
6. Лобанов А.Л., Дианов М.Б., Смирнов И.С. Результаты разработок и использования зоологических информационно-поисковых систем// Информационно-поисковые системы в зоологии и ботанике, Труды Зоологического института РАН, т. 278, Санкт-Петербург, 1999: 54,55.
7. Lobanov A.L., Dianov M.B., Medvedev S.G., Panov V.E., Smirnov I.S. Presentation and use of zoological information in the global Internet network. Trudy Zool. Inst. Ross. Akad. Nauk, 286, 2000: 87-92.
8. Скарлато О.А., Алимов А.Ф., Лобанов А.Л., Умнов А.А. Машинные банки данных - подход к кадастру животного мира// Всесоюзное совещание по проблеме кадастра и учета животного мира, Уфа. 1989. с.56-64.
9. Smirnov I. Working out a databank on marine invertebrates. - In: Abstracts of 8th International Echinoderm Conference. Dijon, September 6 - 10, 1993. P. 111.
10. Смирнов И.С., Лобанов А.Л., Соколов Е.П., Дианов М.Б. Информационно-поисковая система ЗООИНТ для зоологии// Вестник РФФИ, N 2. 1995: 34-36

11. Smirnov I.S., Lobanov A.L., Alimov A.F., Dianov M.B., Medvedev S.G. Development of Information Retrieval Systems for Zoology// Proceedings of the Third International Workshop on Advances in Databases and Information Systems, (ADBIS'96). September 10-13, 1996. V. 2, Moscow, 1996. P. 60-63.
12. Смирнов И.С., Лобанов А.Л., Алимов А.Ф., Дианов М.Б. Развитие интегрированной системы "BIOINT". Третье совещание "Компьютерные базы данных в ботанических исследованиях" С.-Петербург, 20-22 мая 1997 года, Ботанический ин-т РАН. Программа и тезисы докладов. 1997. стр. 44,45.
13. Смирнов И.С., Смирнов А.В. "ОКЕАН" - банк данных по коллекциям морских беспозвоночных Зоологического института РАН // Базы данных и компьютерная графика в зоологических исследованиях, Труды Зоологического института, т. 269, Санкт-Петербург, 1997: 133-135.
14. Смирнов И.С., Лобанов А.Л., Алимов А.Ф., Дианов М.Б. Итоги развития зоологической интегрированной информационной системы ЗООИНТ// Информационно-поисковые системы в зоологии и ботанике, Труды Зоологического института РАН, т. 278, Санкт-Петербург, 1999: 113,114.
15. Смирнов И.С., Рысс А.Ю. Биологические коллекции и базы данных// Информационно-поисковые системы в зоологии и ботанике, Труды Зоологического института РАН, т. 278, Санкт-Петербург, 1999: 30-38.