



**INTERNET
& MODERN
SOCIETY**

**ИНТЕРНЕТ
И СОВРЕМЕННОЕ
ОБЩЕСТВО**

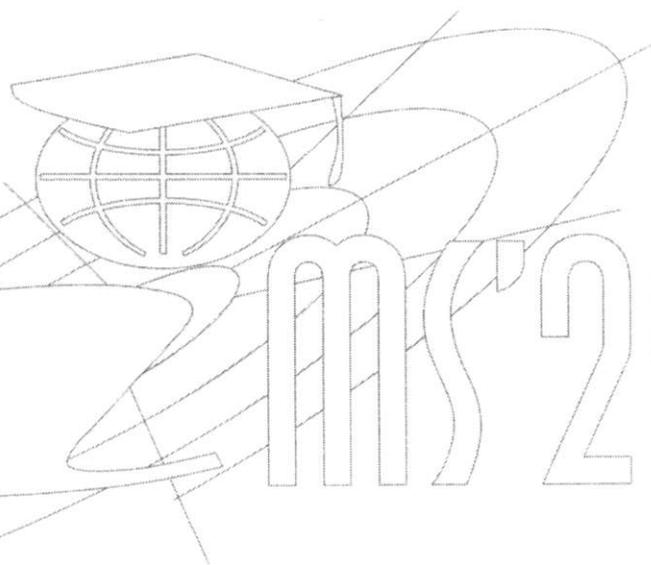


ДЕВЯТАЯ ВСЕРОССИЙСКАЯ ОБЪЕДИНЁННАЯ КОНФЕРЕНЦИЯ

**Всероссийская научная конференция
ГУМАНИТАРНАЯ ИНФОРМАТИКА**

**Всероссийская научная конференция
ЭЛЕКТРОННОЕ ПРАВИТЕЛЬСТВО
В ИНФОРМАЦИОННОМ ОБЩЕСТВЕ**

<http://conf.infosoc.ru>



MS' 2006

САНКТ-ПЕТЕРБУРГ 14 - 18 НОЯБРЯ 2006 ГОДА

Санкт-Петербургский государственный университет

**ИНТЕРНЕТ И СОВРЕМЕННОЕ
ОБЩЕСТВО**

**ТРУДЫ IX ВСЕРОССИЙСКОЙ
ОБЪЕДИНЕННОЙ КОНФЕРЕНЦИИ**
14—16 ноября 2006 г., Санкт-Петербург

Санкт-Петербург 2006

ЗООЛОГИЧЕСКИЕ ЭЛЕКТРОННЫЕ ПУБЛИКАЦИИ: КОЛЛЕКЦИИ И ИДЕНТИФИКАЦИОННЫЕ СИСТЕМЫ

И.С. Смирнов, А.Л. Лобанов, А.Ф. Алимов, О.Н. Пугачев, А.Г. Кирейчук,
В.А. Кривохатский, О.Н. Граничин*, А.Т. Вахитов*

Зоологический институт РАН

* Санкт-Петербургский государственный университет
Санкт-Петербург

ВВЕДЕНИЕ

За прошедшие полвека (достаточно большой срок с учетом бешеного темпа эволюции компьютерных методов и технологий) компьютеризация проникла во все разделы зоологии. Зоологический институт РАН (ЗИН) занимает одно из ведущих мест среди биологических учреждений России в использовании новейших достижений информатики [1,3].

Со времени написания первых обзоров по развитию информационных технологий в зоологии и ботанике [2] прошло более 7 лет. Цель нынешнего доклада — оценить количественные и качественные изменения, которые произошли с компьютерными технологиями и электронными публикациями в Зоологическом институте за прошедший период.

КОЛЛЕКЦИИ

Зоологическая коллекция — это упорядоченное собрание (или набор объектов), представляющих научный или образовательный интерес и хранящееся в специальных учреждениях — естественнонаучных музеях и институтах. Научные коллекции подразделяют на *обзорные* и *исследовательские*. Последние могут содержать законсервированные материалы или живые объекты (культуры). *Мониторинговые* коллекции (периодически повторяющиеся сборы с целью выявления изменений в природных сообществах) — выделяют в отдельный тип исследовательских коллекций. Единицы хранения в коллекциях разнообразны — от банки с большим числом экземпляров одного вида беспозвоночных до части скелета позвоночных [5].

Зоологический институт по числу экземпляров (более 60 миллионов единиц хранения) и уникальности сборов занимает одно из первых мест среди мировых хранилищ коллекций животных.

Интенсивная компьютеризация зоологии ознаменовала новый этап в развитии коллекционного дела. Базы данных (БД) и информационно-поисковые системы (ИПС) по исследовательским коллекциям подразделяют на БД законсервированных образцов и БД коллекций живых организмов [4]. Растет доступность и значимость оцифрованных коллекций [3].

Сайты, посвященные зоологическим коллекциям, исчисляются уже тысячами и десятками тысяч адресов. Появилось большое число ресурсов с обобщающими каталогами ссылок на сайты с данной тематикой. Но при всем при этом, реальных ресурсов, отражающих не только метаданные, но и конкретные коллекции крайне мало.

Отличительной чертой электронных зоологических коллекций является то, что переход от реального коллекционного материала только к электронным изображениям будет медленным. Из-за огромной сложности зоологических объектов, сложившихся традиций и специфики описания новых видов, которое основывается на обязательном существовании типовых или эталонных экземпляров, сохраняется необходимость содержания и реальных, и электронных образцов. Возможно в недалеком будущем, когда молекулярные генетические методы достигнут своего совершенства, а создание электронных копий зоологических объектов превратится в рутинный и недорогой процесс (методы компьютерного моделирования трехмерных объектов, различные виды томографии для изображения внутренней структуры организмов и т.п.), реальные зоологические музеи изживут себя. А до тех пор зоологические коллекции (электронные и реальные) будут сосуществовать и развиваться параллельно. При этом очень важно в зоологических исследованиях, используя реальные коллекции, иметь возможность перепроверить то или иное описание вида по конкретному экземпляру, или даже просто удостовериться в наличии именно данного вида в данной точке. Последнее становится особенно актуальным в свете интенсивного распространения чужеродных, и особенно вредоносных видов (примеры, ставшие классическими: колорадский жук и двустворчатый моллюск-дрейссена).

По-прежнему, одно из первых мест в Интернете занимает электронная коллекция Национального музея естественной истории в Вашингтоне, США (<http://www.mnh.si.edu/rc/>). Появляются такие замечательные сайты, как, например, сайт Амстердамского зоологического музея по редким и вымершим птицам с трехмерными изображениями (<http://ip30.eti.uva.nl/zma3d/>).

Если рассматривать зоологические коллекционные базы данных как наборы электронных документов или образов, то среди разработок и проектов Зоологического института найдется немало заслуживающих внимания, все они представлены сегодня на портале ЗИН РАН.

Это, во-первых, Банк данных по кольцеванию птиц (<http://www.zin.ru/rybachy/ricing0.html>).

Создание компьютерного банка данных Биологической Станции «Рыбачий» (Балтийское море, Россия) началось в 1979 г. Теперь он содержит более чем 2 300 000 записей по окольцованным птицам, 113 000 записей по перелетным птицам и более чем 12 000 записей по возвратам птиц [3].

Во-вторых, ЗООИНТ - ЗООлогическая ИНТЕгрированная система — информационный проект-ветеран, который уже 5 раз поддерживался грантами РФФИ в 1993, 1996, 1999, 2002 и 2005 гг. [3]. Именно от него берут начало многие разработки сотрудников ЗИН в области таксономических баз данных [1, 2, 3]. По мере развития отдельные аспекты концепции ЗООИНТ превратились в самостоятельные направления. Многолетняя работа над коллекционными БД дала целое дерево ответвлений. Центральный ствол представлен интегрированной системой ЗООИНТ. Одна из ветвей ИПС ЗООИНТ дала систему «ОКЕАН», которая коллективно используется и совершенствуется в лабораториях Морских исследований и Ихтиологии. Мощное ответвление представлено информационно-аналитической системой по мировой фауне блох PARHOST, развиваемой С.Г. Медведевым с коллегами [3]. Другая ветвь системы ЗООИНТ развивается с участием А.Ф. Алимова и его коллег. Это ИПС INVADER, которая служит для накопления и анализа данных об организмах-вторженцах [3].

Особое значение базы данных имеют в исследованиях, проводимых в Лаборатории систематики насекомых [3]. Эта Лаборатория имеет самую большую коллекцию в Институте — около 25 млн. единиц хранения, а объекты исследования ее сотрудников относятся к самому большому классу животных — насекомым, включающему более миллиона видов. Поэтому представить себе более или менее полную компьютеризацию сведений об этих коллекциях в настоящее время невозможно. Но разработки баз данных, ориентированных на коллекционные сведения, ведутся, и примером такого проекта может служить информационная система «ZInsecta», разрабатываемая В.А. Кривоухатским с коллегами и включающая разнообразные авторские БД по иерархической классификации и коллекциям насекомых [3].

Информационно-поисковая система «Океан» представляет собой первый в ЗИН распределенный банк данных. В Интернете ИПС «Океан» развивается на сайтах арктической программы по биоразнообразию ArcOD и антарктической программы CAML. К вводу данных, осуществляемому по-прежнему в СУБД Fox-Pro for Windows, прибавились работы по миграции данных из формата Fox-Pro в формат MS SQL Server с одновременной тщательной унификацией полей и данных, созданием словарных таблиц. Организуется вывод данных по запросам в веб-интерфейсе на базе ASP-технологии пока в пределах институтского Интранета.

ИДЕНТИФИКАЦИОННЫЕ СИСТЕМЫ

В обзоре 2001 г. авторы данного доклада написали, что «не наблюдается принципиальных новшеств и оригинальных теоретических находок» в компьютерных интерактивных системах [2]. Данное высказывание остается в силе и на сегодняшний момент, за исключением вторжения Интернет-технологий в возможности использования определительных систем в глобальной сети. Наиболее просто конструируются при этом одновходовые ключи, легко реализуемые средствами гипертекста. Даже используя только минимальный набор тэгов языка HTML можно построить достаточно удобный и эффектный ключ. Использование Java-скриптов и апплетов позволяет дополнить такой ключ очень полезными возможностями. К сожалению, еще мало попыток создания в Интернете настоящего многовходового ключа, работающего со стандартной базой данных, например, через интерфейс CGI или ODBC. Но быстрое развитие инструментальных средств глобальной сети Интернет создает новые возможности /утя новых интересных реализаций многовходовых ключей, существенно повышающих эффективность диагностики и использования таксономических и коллекционных баз данных.

Значительное место среди компьютерных разработок ЗИНа занимает диалоговая компьютерная диагностическая система VIKEY/PICKEY/WEBKEY-X. Биологическая компьютерная диагностическая система VIKey (Biological Identification Keys) предназначена для автоматизации процессов работы с диагностической информацией о биологических таксонах и также активно использует БД. Эта система является развитием идеи интерактивного использования информации по биоразнообразию и коллекциям [3].

С помощью этой системы был подготовлен определитель жесткокрылых насекомых, опубликованный в Германии в виде компьютерного компакт-диска. Созданы первые версии определителей иглокожих, в частности — арктических (в MS DOS) и антарктических (в MS Windows) офиур. А.Ю. Рыссом созданы ключи для фитопатогенных нематод [3]. Подготовка данных для этой системы является кропотливым и отнимающим много времени трудом, который, однако, окупается удобством и надежностью создаваемых компьютерных определителей, с которыми ни в коей мере не могут конкурировать традиционные «бумажные» издания.

В 2005 г. А.Г. Кирейчук с коллегами получил грант РФФИ на разработку интерактивной диагностической программы для использования в сети Интернет (РФФИ N 05-07-90179). Это дало возможность подключить к работе университетских специалистов в области сетевых компьютерных систем на основе баз данных — О.Н. Граничина и А.Т. Вахитова. Так родился проект WebKey-X (<http://www.zin/projects/webkey-x/index.html>): «Разработка интерактивной, ориентированной на работу в сети Интернет и использующей технологию баз данных, программы для определения биологических объектов».

Существующие в Интернете интерактивные определители или не приспособлены для работы в сети или приемлемы только для работы с нестандартными базами данных. В целом они не имеют математически обоснованного механизма выбора ценных в диагностическом смысле признаков. Разрабатываемый проект WebKey-X призван преодолеть эти трудности и создать программную основу для эффективного использования многоходовых политомических определителей в Интернете, а также обеспечить систематиков, фаунистов, экологов и специалистов по прикладным областям биологии легко доступной таксономической информацией в сети Интернет, позволяющей быстро определять эти группы животных. Разработанное программное обеспечение будет использовано для составления определителей, и том числе и по группам, базы данных по которым в настоящее время размещены на других сайтах веб-портала ЗИН РАН, а в дальнейшем предполагается и создание отчуждаемых копий системы и определителей на DVD-дисках.

Интерактивная программа для определения биологических объектов WebKey-X создана на основе технологии Java 2 Enterprise Edition (J2EE), которая используется для организации и поддержки в сети Интернет информационных систем, как малых, так и более объемных, нежели WebKey-X. При выборе технологии также играли роль удобство разработки приложений и переносимость. Альтернативой J2EE является организация системы на основе продуктов фирмы Microsoft C# и .NET Framework. Технология J2EE была выбрана как перспективная, удобная для расширения, не уступающая .NET по простоте поддержки и использованию ресурсов. И, что немаловажно, для работы создаваемых приложений достаточно использовать бесплатно и свободно распространяемые программные продукты (Open Source).

Созданная программа отличается своей простотой и, тем не менее, реализует все основные функции, определенные техническим заданием на первый год реализации проекта. Выгодным отличием от многих представленных в Интернете определителей является многоходовость — возможность начать идентификацию таксона с любого удобного пользователю признака. Тем не менее, признаки предлагаются не в случайной последовательности: они сортируются в порядке значимости для идентификации на каждом шаге. Метрика, с помощью которой определяется предпочтительность признака (его «диагностическая ценность») оформлена в виде отдельного блока программы и без труда может быть заменена. Это делает достаточно легким процесс поиска наилучшей метрики. Важным достоинством программы является то, что все данные конкретного ключа хранятся в базе данных стандартного формата: DBF (DataBase) или MDB (Access). В настоящее время система работает с базой данных по 130 видам жуков и обеспечивает приемлемую скорость ответа на каждый запрос и время идентификации какого-либо экземпляра. Реализованы функции протокола идентификации, изменения состояния (или нескольких состояний) уже использованного признака; выдается описание идентифицированного вида в конце работы и список допустимых вариантов видов на любом шаге. Алгоритмически оптимизировать остается только процесс оценки предпочтительности признаков, не забывая о сохранении хорошей скорости работы системы. Недопустимость переборных решений даже для 130 видов говорит о невозможности создания строго оптимального механизма сортировки, поэтому следует ставить вопрос о наиболее близком к оптимальному решению в условиях ограничений на время отклика системы. Также важным для потенциальных пользователей является понятный интерфейс и наличие необходимых функций в системе.

Существует и другая, более простая Интернет-технология, когда специалисту через глобальную сеть пользователь высылает фотографии экземпляров, интересующих его животных (сейчас цифровая камера перестала быть редкостью), а систематик по изображениям определяет тот или иной вид (<http://www.zin.ru/Animalia/Coleoptera/rus/tegroder.htm>; <http://www.zin.ru/Animalia/Coieoptera/rus/chechet2.htm>).

Разными путями цифровые методики и направления проникают в различные области, не только собственно зоологических исследований, но и представления их результатов как в Интранете, так и в Интернете.

Применение новейших компьютерных технологий и, особенно, Интернет-технологий в Зоологическом институте РАН приобретает все более широкие масштабы и, несомненно, должно способствовать повышению эффективности и достижению высокого научного уровня проводимых исследований.

Работа по теме осуществляется при поддержке грантов РФФИ N 05-07-90179-в, 05-07-90354-в и 06-04-08020-офи, проекта «Антарктика» (ФЦП «Мировой океан»), программы «Биоразнообразия», проектов ArcOD и CAML.

ЛИТЕРАТУРА

1. *Алимов А.Ф.* Компьютеризация биологического учреждения (на примере Зоологического института РАН) // Рысс А.Ю., Смирнов И.С. (ред.). Информационно-поисковые системы в зоологии и ботанике: Труды Зоологического института РАН. Т. 278. Санкт-Петербург, 1999. С. 7 - 8.
2. *Alimov A.F., Smirnov I.S., Ryss A. Y., Dianov M.B., Lobanov A.L., Golikov A.A.* Modern biological electronic publications: collections, identification systems and databases // Information technology in biodiversity research. 2001. Pp. 5-19.
3. *Лобанов А.Л., Смирнов И.С.* Место и роль информационных технологий в исследованиях Зоологического института РАН // Фундаментальные зоологические исследования: Теория и методы. (По материалам Международной конференции «Юбилейные чтения, посвященные 170-летию Зоологического института РАН», проходившей 23-25 октября 2002 г.). М.-СПб.: Товарищество научных изданий КМК. 2004. С. 283 -3 18 (резюме на англ. яз.).
4. *Смирнов И.С., Рысс А.Ю.* Биологические коллекции и базы данных // Рысс А.Ю., Смирнов И.С. (ред.). Информационно-поисковые системы в зоологии и ботанике // Труды Зоологического института РАН. Т. 278. Санкт-Петербург, 1999. С. 30 - 38.
5. *Соколов Е.П., Смирнов И.С., Лобанов А.Л.* Интегрированная система ZOOINT для ведения и использования зоологических баз данных // Степаньянц С.Д., Лобанов А.Л., Дианов М.Б. (ред.). Базы данных и компьютерная графика в зоологических исследованиях // Труды Зоологического института РАН, т. 269. 1997. С. 136-144.