

ПРОБЛЕМЫ И РЕЗУЛЬТАТЫ РАЗРАБОТКИ ПРОЕКТА ИСИБР «ИНФОРМАЦИОННАЯ СИСТЕМА ПО БИОРАЗНООБРАЗИЮ ЖИВОТНЫХ РОССИИ»

Р. Г. Халиков, М. Б. Дианов, А. Л. Лобанов, И.С. Смирнов, А. А. Голиков

Разработка таксономических баз данных (БД) и информационных систем (ИС) началась в Зоологическом институте РАН (ЗИН) в 1986 г. (Лобанов, Смирнов, 2004).

Биологические базы данных опираются на таксономические таблицы, в которых должны полно отражаться сложные многоуровневые иерархические схемы классификации таксонов. Отражение иерархий в плоских таблицах реляционных баз данных является нетривиальной задачей. Несколько способов решения этой проблемы известны в теории информационных систем (Van Tulder, 2003; Шеваев, 2004). В ЗИНе большинство этих способов было разработано независимо в ходе многолетних работ по созданию зоологических баз данных (Лобанов, Смирнов, 2004). Все эти разработки были отражены в стандарте ZOOCOD, многократно описанном в ряде публикаций (Лобанов, Зайцев, 1991; Лобанов, Смирнов, 1997; Лобанов и др., 1999; Lobanov et al., 2003).

Традиционный подход с применением СУБД FoxPro в последнее время стал создавать ряд проблем: персональная СУБД; проблемы многопользовательского доступа, производительности, масштабируемости и расширяемости; структура таблиц: систематический порядок определяется физическим порядком записей; синонимичные таксоны располагаются строго под валидным таксоном; необходимость использования расчетного поля HERCOD для удобства отображения многоуровневой иерархии в форме таксономического древа; программирование в среде FoxPro: проблема «одного разработчика», закрытая система; проблемы совместимости баз данных FoxPro с современными версиями управляющих элементов и программных компонентов; проблема кодировок.

Совокупность недостатков заставили обратиться к более совершенным программным продуктам, а именно к программе MS SQL Server, для которой можно отметить следующие ключевые преимущества: клиент-серверная СУБД корпоративного уровня: многопользовательский доступ с разделяемыми правами, производительность, масштабируемость и расширяемость; промышленный стандарт структуры данных, построения запросов, средств импорта и экспорта данных, поддержка производителя и совместимость с новыми версиями программных компонентов, поддержка юникода; мощные встроенные средства программирования, расчета значений полей и автоматического заполнения полей, проверки и обеспечения целостности данных; средства централизованного управления элементами баз данных, пользователями и правами доступа, развитые средства обеспечения безопасности и резервного копирования.

Преимущества, которые просматриваются в рамках проекта ZooDiv: клиент-серверная информационная система: разделение данных (таблицы в СУБД – серверная часть) и представления данных (пользовательский интерфейс – клиентская часть); использование единого хранилища унифицированных таксономических данных с разделяемым доступом для различных задач и проектов, как внутри сети ЗИН РАН, так и в публичном доступе на веб-портале; импорт имеющихся разрозненных таксономических данных в единый классификатор, предоставление широкому кругу специалистов средств для внесения исправлений и дополнений, создание удобного пользовательского интерфейса для заполнения классификатора новыми данными.

Реализация наработок в рамках проекта ZooDiv привела к возникновению новых проблем и постановке новых задач, диктуемых длительностью разработки структуры системы и алгоритмов обработки данных: ограниченные ресурсы разработчиков: «делать, чтобы не переделывать»; специфичность таксономических данных и методов работы с ними: невозможность прямого использования готовых технических решений и привлечения разработок сторонних специалистов. Огромные объемы данных обусловили необходимость проведения тестов для оценки производительности и масштабируемости реализуемых алгоритмов; создание открытой системы предопределило использование стандартных средств программирования - MS SQL Server (серверная часть) и универсального веб-интерфейса (клиентская часть – технология ASP; кросс-платформенный скриптовый язык JavaScript, принципиально одинаковый в «серверной» и «клиентской» частях кода ASP-страниц). Потребовались пересмотр стандарта ZOOCOD и поддержка строгой реляционной модели (вспомогательные таблицы, полная нормализация).

Работа выполнена при поддержке РФФИ (проекты № 05-07-90179-в, 05-07-90354-в и 06-04-08020), программ "Антарктика" (ФЦП "Мировой океан") и "Биоразнообразие".

ЛИТЕРАТУРА:

1. Лобанов А.Л., Зайцев М.В. Создание компьютерных баз данных по систематике млекопитающих на основе классификатора названий животных "ZOOCOD" // Вопросы систематики, фаунистики и палеонтологии мелких млекопитающих (Труды Зоологического института РАН, т. 243) . СПб, 1991, с. 180-198.

2. 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.
3. Лобанов А.Л., Смирнов И.С. Место и роль информационных технологий в исследованиях Зоологического института РАН // Фундаментальные зоологические исследования. Теория и методы. М. - СПб. Изд-во КМК. 2004. С. 283-318.
4. Лобанов А.Л., Смирнов И.С., Дианов М.Б. ZOOCOD - концепция представления зоологических иерархических классификаций в реляционных базах данных // Информационно-поисковые системы в зоологии и ботанике (Тезисы международного симпозиума, май 1999). Труды Зоологического института РАН. 1999. Vol. 278. P. 66.
5. Щеваев П.А. Способы хранения иерархических структур в реляционных базах данных // Новые информационные технологии и системы: Труды VI Международной научно-технической конференции. Часть 2. Пенза, ПГУ, 2004. С. 226-233.
6. Lobanov A.L., A.Yu. Ryss, I.S. Smirnov. A modern state of the ZOOCOD concept // Information Systems on Biodiversity of Species & Ecosystems. Scientific program & abstracts, SPb. 2003. pp. 7, 8
7. VanTulder Gijs, 2003. Storing hierarchical data in a DB: <http://www.sitepoint.com/print/hierarchical-data-database>.