

ПРОЕКТИРОВАНИЕ БАЗЫ ДАННЫХ "ПЛАНКТОН ЧЕРНОГО МОРЯ" С ИСПОЛЬЗОВАНИЕМ ER-ДИАГРАММ

Предлагается проект базы данных "Планктон Черного моря", предназначенной для хранения первичных данных, полученных в результате обработки проб. Необходимые для проектирования данные были собраны в результате изучения отчетов по рейсам научно-исследовательских судов, а также проведения "интервью" со специалистами – гидробиологами ИнБЮМ НАНУ. В качестве концептуальной модели при проектировании использовалась диаграмма "сущность-связь" (ER – диаграмма).

В гидробиологии используется огромный объем информации, которую необходимо хранить и обрабатывать. Наиболее эффективно применять для этих целей системы баз данных. Хранение информации, организованной в виде базы данных, дает следующие преимущества:

- 1) обеспечивается наиболее надежное хранение при наименьших затратах, минимальная избыточность данных и быстрый поиск требуемой информации;
- 2) информация из базы данных может использоваться любой обрабатывающей программой;
- 3) возможность использования компьютерных сетей: компьютерные сети (в частности Internet) позволяют исследователю получить необходимые данные из любого сервера сети или предоставить свои данные по запросу коллег.

Гидробиологические данные образуют несколько основных уровней: а) самый верхний – уровень обобщения (информация, содержащаяся в учебниках и фундаментальных монографиях); б) содержимое определителей и руководств по систематике; в) первичные данные, полученные в результате обработки проб. Особенный интерес, с точки зрения исследователей, представляет последний из вышеперечисленных уровней [1]. В статье представлен пример проектирования базы данных для хранения первичных данных по отдельным таксонам планктона Черного моря.

Первым этапом при проектировании баз данных является анализ предметной области. Задачи этого этапа заключаются, во-первых, в получении проектировщиком общего представления о предметной области; во-вторых, в сборе элементов данных, которые требуется отобразить в базе данных; в-третьих, в формулировании ограничений целостности, которым должны удовлетворять данные будущей базы; в-четвертых, в формировании набора основных транзакций и документов, возможность выдачи которых должна обеспечить проектируемая база данных.

Для решения вышеперечисленных задач использовались отчеты по рейсам научно-исследовательских судов ИнБЮМ и экспертные знания специалистов-гидробиологов, полученные в результате проведения «интервью». В результате анализа была получена следующая, интересующая проектировщика информация. Судно, обходя поочередно несколько районов исследований, подходит к станциям, на которых исследователи берут пробы. Каждое взятие проб отличается датой, временем, глубиной и орудием лова. Элементы данных, которые представляют интерес: пробы, измеряемые в каждой пробе параметры, главным образом численность таксономических групп (набор измеряемых параметров может меняться в зависимости от требований организации – исследователя; при внедрении настоящей базы данных на конкретном предприятии заказчик может подкорректировать необходимый набор параметров), таксоны, их классификационное положение, средний размер особи данного таксона в пробе, средняя калорийность биомассы, горизонт (верхняя и нижняя границы в метрах), район исследования, автор исследования, организация, к которой этот автор принадлежит, рейс, судно, характеристики судна, станции, координаты станций (широта, долгота), время взятия пробы, орудие лова, биомасса в пробе, глубина станции, дата (число, месяц, год), время суток (часы, минуты).

Выявленные ограничения целостности: 1) на каждом судне могут находиться исследователи, принадлежащие к различным организациям; 2) каждая проба принадлежит только одному исследователю; 3) на каждой станции пробы могут брать все заинтересованные исследователи; 4) в каждом районе исследования может находиться несколько станций; 5) каждая станция характеризуется определенными границами горизонтов; 6) на каждой станции пробы могут быть взяты на различной глубине в пределах границ горизонтов.

К основным транзакциям относятся: 1) добавить информацию о пробе; 2) добавить информацию о таксоне; 3) добавить информацию об исследователе; 4) добавить информацию о судне; 5) изменить данные о станции; 6) изменить данные о судне; 7) изменить данные о рейсе; 8) изменить данные об организации; 9) изменить данные об исследователе; 10) изменить данные о таксоне; 11) удалить информацию о пробе; 12) удалить информацию о таксоне; 13) удалить информацию о судне; 14) удалить информацию о рейсе; 15) удалить информацию об исследователе.

Должна быть обеспечена возможность выдачи следующих документов: 1) отчеты о рейсах, включающие название судна, районы исследования, данные об организациях, участвующих в экспедиции; 2) отчеты по пробам, содержащие основные измеряемые в каждой пробе параметры; 3) отчеты исследователей, включающие статистические данные о собранных пробах (количество проб, место взятия, время взятия, орудие лова и т.д.); 4) справка о станциях (по районам исследования), включает сведения о границах горизонтов, координатах; 5) статистические данные об исследованных в пробах таксонах (численность таксона в пробе, средний вес особи и т.д.).

Следующим этапом проектирования является построение концептуальной модели предметной области. В качестве концептуального представления данных будем использовать диаграмму «сущность-связь» (ER-диаграмму). Для построения ER-диаграммы сначала следует выделить среди собранных элементов данных сущности. В качестве сущностей выберем: ПРОБА, ТАКСОН, РЕЙС, СТАНЦИЯ. Далее установим связи между сущностями. Каждая проба берется на какой-то станции; следовательно, между сущностями ПРОБА и СТАНЦИЯ существует связь, которую назовем ВЗЯТА. Каждая проба включает несколько таксонов; следовательно, между ПРОБА и ТАКСОН есть связь, назовем ее ВКЛЮЧАЕТ. Судно, совершающее рейс, обходит несколько станций; значит, между сущностями РЕЙС и СТАНЦИЯ существует связь, присвоим ей имя СОСТОИТ. Теперь необходимо распределить оставшиеся элементы данных в качестве атрибутов по сущностям или связям. Принцип распределения состоит в том, что относим к каждой из сущностей те атрибуты, которые однозначно являются ее характеристикой и отнесение их к другой сущности нарушит семантику сущности или атрибута. Например, атрибуты *название таксона*, *место в классификации* отнесем к сущности ТАКСОН, а *номер станции*, *координаты станции* к сущности СТАНЦИЯ. Но не все атрибуты можно так тривиально распределить. Например, атрибут *численность таксона в пробе* нельзя отнести к сущности ПРОБА, т.к. в каждой пробе могут присутствовать представители нескольких таксонов, поэтому включение этого атрибута создаст избыточность данных в отношении ПРОБА, поскольку для каждого значения численности придется повторять все значения остальных атрибутов этого отношения. Аналогично, нельзя отнести этот атрибут к сущности ТАКСОН, поскольку в каждой пробе будет свое значение этого атрибута, следовательно, включение его в отношении ТАКСОН также приведет к избыточности. Выход из сложившегося затруднения состоит в том, чтобы отнести этот атрибут к связи ВКЛЮЧАЕТ. Поскольку отношение-связь включает в качестве атрибутов ключи сущностей ТАКСОН и ПРОБА, то мы сможем получить информацию о том, какова численность конкретного таксона в конкретной пробе, избежав избыточности данных. Рассуждая подобным образом, распределим остальные атрибуты. Построенная диаграмма представлена на рис. 1. Степени связи и классы принадлежности сущности в связи обозначены на диаграмме в форме, принятой в [2]. Подчеркнутые на диаграмме атрибуты выбраны в качестве первичных ключей отношений.

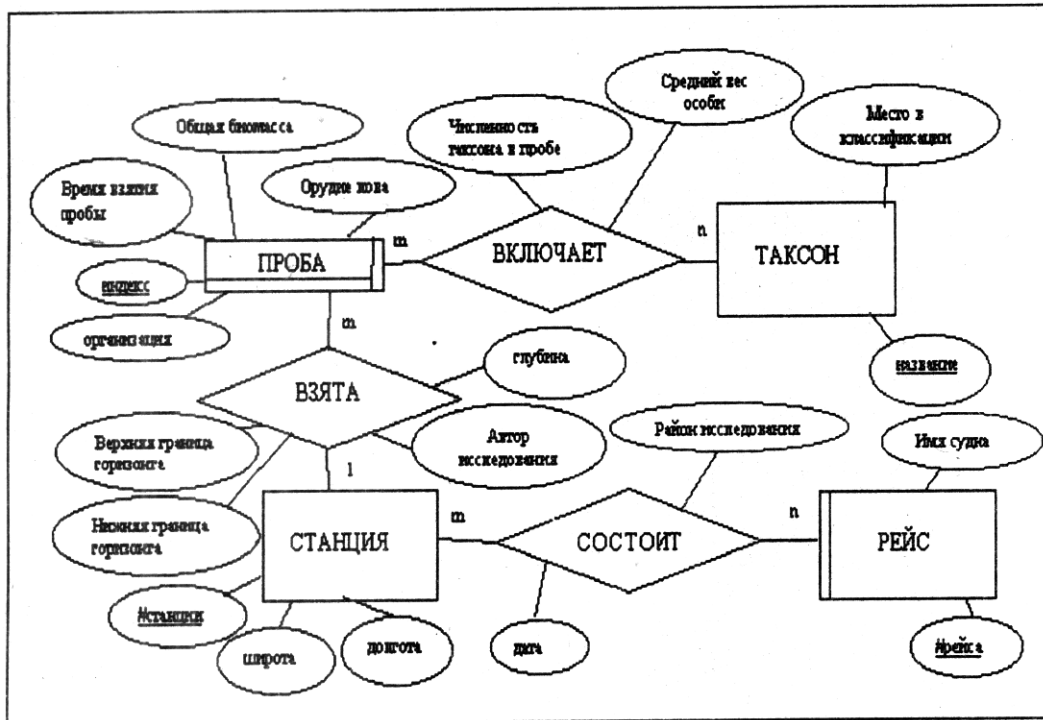


Рис.1 ER- диаграмма
Fig.1 ER- diagram

Одним из достоинств ER- диаграмм является хорошо формализованный переход от диаграммы к предварительным отношениям. В результате перехода получены следующие предварительные отношения:

ПРОБА (индекс, время взятия пробы, общая биомасса, орудие лова, организация-исследователь, №станции, глубина, автор исследования).

СТАНЦИЯ (№станции, долгота, широта, нижняя граница горизонта, верхняя граница горизонта).

ТАКСОН (название, место в классификации).

ВКЛЮЧАЕТ (индекс, название, численность таксона в пробе, средний вес особи).

РЕЙС (№рейса, имя судна).

СОСТОИТ (№станции, №рейса, район исследований, дата).

Для проверки, являются ли предварительные отношения нормализованными (т.е. находятся в нормальной форме Бойса-Кодда (НФБК)) или нет [3], необходимо найти функциональные зависимости в каждом отношении.

1) Отношение ПРОБА:

индекс $\rightarrow$ время взятия пробы, общая биомасса, орудие лова, организация-исследователь, глубина, автор исследования, №станции;

единственным возможным ключом данного отношения является атрибут индекс и все непервичные атрибуты функционально зависят от ключа, следовательно, отношение находится в НФБК.

2) Отношение СТАНЦИЯ:

№станции $\rightarrow$ долгота, широта, нижняя граница горизонта, верхняя граница горизонта;

долгота широта $\rightarrow$ №станции, нижняя граница горизонта, верхняя граница горизонта;

возможными ключами отношения являются: №станции и долгота-широта, значит, отношение также находится в НФБК.

3) Отношение ТАКСОН:

название → место в классификации;

место в классификации → название;

оба атрибута данного отношения являются возможными ключами, следовательно, отношение в НФБК.

4) Отношение ВКЛЮЧАЕТ:

индекс название → численность таксона в пробе, средний вес особи;

возможным ключом является совокупность атрибутов индекс-название, т.е. отношение также в НФБК.

5) Отношение РЕЙС:

№ рейса → имя судна;

отношение в НФБК.

6) Отношение СОСТОИТ:

Местанции № рейса → район исследования, дата;

отношение находится в НФБК.

Таким образом, мы убедились, что каждое из полученных отношений находится в НФБК, значит, схема базы данных также находится в НФБК.

На следующем этапе на основании собранного набора основных документов и транзакций осуществляется тестирование схемы базы данных, т.е. анализируется возможность реализации каждой из транзакций в пределах полученной схемы, а также проверяется возможность выдачи требуемых документов по данным, хранящимся в базе. Тестирование спроектированной схемы подтверждает ее полноту и достаточность. Полученная схема может быть реализована в среде любой СУБД, поддерживающей реляционную модель данных.

Спроектированная база данных предназначена для хранения первичных данных о планктоне, но возможно построение баз данных, хранящих данные из фундаментальных источников и содержимое руководств по систематике и определителей. Все эти базы данных могут быть объединены в единую систему и отображать всю известную о планктоне информацию. Тем самым будет решаться задача "инвентаризации" планктона в заданном регионе. При современном уровне компьютерной техники, позволяющей хранить не только текстовую информацию, но и рисунки, наличие подобной системы значительно облегчит гидробиологические исследования и упростит обмен данными между исследователями.

1. Михайловский Г.Е. Описание и оценка состояний планктонных сообществ. - М.: Наука, 1988. - 211 с.
2. Джексон Г. Проектирование реляционных баз данных для использования с микроэвм. -- М.: Мир, 1991. -- 248 с.
3. Дейт К. Введение в системы баз данных. -- М.: ЮНИТИ, 1997. -- 681 с.

Институт биологии южных морей НАНУ,
г. Севастополь

Получено 15.08.2000

J. G. SHATALOVA

THE DATABASE DESIGN USING ER – DIAGRAM

Summary

The procedure of the database construction for accumulation of the information about plancton is presented. The entity – relationship diagram is used as an instrument of that procedure. The applications of such database are indicated.