

ПРОЕКТИРОВАНИЕ БАЗЫ ДАННЫХ СИСТЕМЫ МОДЕЛЕЙ ОПТИМАЛЬНОГО ПЛАНИРОВАНИЯ СЕЛЬСКОХОЗЯЙСТВЕННОГО ПРОИЗВОДСТВА

В.И. Векленко, Н.Н. Петренко, В.Ю. Пашков, О.В. Ерёменко

Аннотация. Разработана методика построения документальной БД для компьютерной системы оптимального планирования сельскохозяйственного производства и выполнено инфологическое описание модели БД для конкретного агропроизводственного объекта – СХПК «Новая жизнь» Беловского района Курской области.

Ключевые слова: концептуальное проектирование, база данных (БД), модель БД, фактографические и документальные БД, инфологическая модель БД, архитектура системы управления БД (СУБД).

Наиболее простой способ представления предметных областей в БД реализуется поэтапно:

– фиксацией логической точки зрения на данные (т. е. данные рассматриваются независимо от особенностей их хранения и поиска в конкретной вычислительной среде);

– определением физического представления данных с учетом выбранных структур хранения данных и архитектуры ЭВМ [1].

Абстрагированное описание предметной области с фиксированной (логической) точки зрения составляет концептуальную схему. Соответственно, систематизация понятий и связей предметной области называется логическим или концептуальным проектированием. Модель (представление логической точки зрения), используемая при абстрагировании — совокупность функциональных характеристик объектов и особенностей представления информации (представляет собой модель данных [5].

Отображение концептуальной схемы на физическом уровне может быть названа внутренней схемой. Тогда отражение взгляда (точки зрения) отдельного пользователя на концептуальную схему (как вариант восприятия предметной области) будет представлять собой внешнюю схему. Внешняя схема использует те же абстрактные категории, что и концептуальная, а на практике соответствует логической организации данных в прикладной программе.

Таким образом, любое состояние объекта характеризуется совокупностью актуализированных атрибутов (имеющих некоторое из значений в этот момент времени), которые фиксируются на некотором материальном носителе в виде записи — совокупности (группы) formalizованных элементов данных (значений атрибутов, представленных в том или ином формате). Кроме того, в контексте задачи хранения и поиска можно утверждать, что значение атрибута идентифицирует объект: использование значения в качестве поискового признака позволяет реализовать простой критерий отбора по условию сравнения.

Предложенная выше схема представляет атрибутивный способ идентификации содержания объекта. Она является достаточно естественной для данных, имеющих фактографическую природу и описывающих обычно материальные объекты. Информацию, представляемую такого рода данными, называют хорошо структурированной. Здесь важно отметить, что структурированность относится не только к форме представления данных (формат, способ хранения), но и к способу интерпретации значения пользователем: значение параметра не только представлено в предопределенной форме, но и обычно сопровождается указанием размерности величины, что позволяет пользователю понимать ее смысл без дополнительных комментариев. Таким образом, фактографические данные предполагают возможность их непосредственной интерпретации.

Однако атрибутивный способ практически не подходит для идентификации слабо структурированной информации, связанной с объектами, имеющими обычно идеальную (умозрительную) природу. Такие объекты зачастую определяются логически и опосредованно — через другие объекты. Для описания таких объектов используются естественные или искусственные языки. Соответственно, для понимания смысла пользователю необходимо использовать соответствующие правила языка, и, более того, часто необходимо уже располагать некоторой информацией, позволяющей идентифицировать и связать получаемую информацию с наличным знанием. Таким образом, процесс интерпретации такого рода данных имеет опосредованный характер и требует использования дополнительной информации, причем такая, которая не обязательно присутствует в формализованном виде в базе данных. Указанное различие нашло отражение в традиционном разделении баз данных на фактографические и документальные [3].

Пользователю необходимо иметь возможность обращаться к отдельным, нужным ему записям (описаниям объектов) или отдельным элементам данных. В зависимости от уровня программного обеспечения это можно достичь следующими способами:

– Задать машинный адрес данных и в соответствии с физическим форматом записи прочитать значение. Это случай, когда пользователь должен быть «навигатором» и иметь соответствующую подготовку;

– Сообщить системе имя записи или элемента данных, которые он хочет получить, и возможно, организацию набора данных. В этом случае система сама произведет выборку (по предыдущей схеме), но для этого нужна вспомогательная информация о структуре данных и организации набора. Такая информация по существу будет избыточной по отношению к объекту, однако общение с базой данных не будет требовать от пользователя специальных знаний и позволит выполнить размещение данных автоматически.

В качестве ключа, обеспечивающего доступ к записи, можно использовать идентификатор – отдельный элемент данных.

Если бы назначением базы данных было только хранение и поиск данных в массивах записей, то структура системы и самой базы была бы простой. Причина сложности в том, что практически любой объект характеризуется не только параметрами-величинами, но и взаимосвязями частей или состояний. Есть различия и в характере взаимосвязей между объектами предметной области: одни объекты могут использоваться только как характеристики остальных объектов, другие - независимы и имеют самостоятельное значение.

Кроме того, сам по себе отдельный элемент данных (его значение) ничего не представляет. Он приобретает смысл только тогда, когда связан с атрибутом (природой значения, что позволит интерпретировать значение) и другими элементами данных.

Поэтому физическому размещению данных (и, соответственно, определению структуры физической записи) должно предшествовать описание логической структуры предметной области — построение модели соответствующего фрагмента реального мира, выделяющей только те объекты, которые будут интересны будущим пользователям, и представленные только теми параметрами, которые будут значимы при решении прикладных задач. Такая модель будет иметь очень мало физического сходства с реальностью, но будет полезна как представление пользователя о реальном

мире. Причем это представление будет задаваться (описываться) удобными для пользователя средствами в не адекватной человеку жесткой вычислительной среде с двоичной логикой и числовым представлением информации.

Таким образом, прежде чем описывать физическую реализацию объектов и связей между ними, необходимо определить:

- 1) способ, с помощью которого внешние пользователи представляют (описывают) объекты и связи;
- 2) форму и методы компьютерного представления элементов данных и взаимосвязей;
- 3) средства, обеспечивающие взаимно однозначные преобразования внешнего и компьютерного представлений.

Такой подход является компромиссом, свойственным языкам программирования: за счет предварительно определяемого множества абстракций, общих для большинства задач обработки данных, обеспечивается возможность построения надежных программ обработки. Пользователь, используя ограниченное множество формальных, но достаточно знакомых понятий, выделяя сущности и связи, описывает объекты и связи предметной области; в системе автоматизации проектирования БД на основе таких типовых абстрактных понятий, как на числа, множества, последовательности, агрегаты, определяются соответствующие информационные структуры. Система управления данными, используя двоичные формы представления типизированных данных, обеспечивает эффективные процедуры хранения и обработки данных.

Именно введение промежуточного уровня абстракции позволяет иметь раздельное описание логического и физического представлений, освобождает пользователя от необходимости работы в деталях компьютерного представления и обработки. Однако это возможно лишь в том случае, если выбрана наиболее эффективная форма автоматизации работы с данными, соответствующая определенной ситуации. Эффективность же имеет определенные пределы: чем ближе система абстракций к особенностям вычислительной среды, тем выше эффективность работы системы, а вынужденная «специализация» абстракций увеличивает вероятность того, что они станут непригодными для других применений.

Модель данных должна, так или иначе, дать основу для описания данных и манипулирования данными, а также дать средства анализа и синтеза структур данных.

Состояние предметной среды с точки зрения описания целесообразно условно разделить на два полярных случая.

1. Предметная среда характеризуется сравнительно небольшим количеством типов отношений, но каждое отношение само есть большое множество. Эти отношения сравнительно устойчивы, а изменений в пределах каждого множества существенно меньше мощности самого отношения. Например, отношение «вхождения» элементов изделий, содержащееся в конструкторских спецификациях, для среднего предприятия содержит сотни тысяч записей. В этом случае, задав схемы отношений и ориентировочные значения их мощностей, можно достаточно полно представить структуру и масштаб предметной среды.

2. Для предметной среды характерно большое число типов отношений между объектами, но каждое отношение есть множество сравнительно малой мощности. При этом мощность потока изменений для отношений сравнима с мощностью самих отношений.

Первый случай характерен для отображения процессов на уровне автоматизированных систем управления предприятиями. Современные системы управления

базами данных наиболее эффективны именно в подобном случае при отображении статического состояния предметной среды. Обычно при этом речь идет о целых классах объектов, например, о деталях данного типа и не отображается состояние каждой конкретной детали.

Второй случай характерен для описания производственного технологического процесса с учетом временных и пространственных факторов нахождения конкретных объектов.

Если в первом случае говорят о реляционной, иерархической или сетевой моделях данных, то во втором - о семантических сетях и фреймах.

Основное отличие этих методов заключается в том, что первые задают четкую схему (так называемую схему базы данных), в рамках которой и отображается предметная область. Подобное построение по сути своей является довольно статичным, требует априорного знания типов отношений, в которых может находиться объект, однако зафиксированная схема базы данных позволяет довольно эффективно организовать поиск необходимой информации. Во втором случае предметная среда отображается (по крайней мере, на уровне модели) в виде однородной сети, любые изменения которой по вводу как новых классов объектов, так и новых типов отношений, не связаны с какими-либо структурными преобразованиями сети. В силу большого количества типов отношений манипулирование подобной «элементарной» информацией достаточно затруднено, поэтому для данного случая характерно введение большого количества общих понятий (и соответствующих им отношений), что упрощает работу с таким представлением.

В контексте компьютерного представления модель данных может быть использована следующим образом:

- как средство спецификации типов данных и их организации, разрешенных в конкретной БД;
- как основа разработки общей методологии построения баз данных;
- как основа минимизации влияния эволюции баз данных на уже существующие прикладные программы и работу пользователей;
- как основа разработки семейства языков запросов и языков манипулирования данными;
- как основа архитектуры СУБД;
- как основа изучения динамических свойств различных организаций данных.

Таким образом, модель данных — это базовый инструментарий, обеспечивающий на формальном абстрактном уровне конкретные способы представления объектов и связей [5].

Модель базы данных охватывает более широкий спектр понятий. Основное назначение модели базы данных состоит в том, чтобы:

- определить ясную границу между логическим и физическим аспектами управления базой данных (независимость данных);
- обеспечить пользователям и разработчикам БД возможность и средства общего понимания смысла данных (коммуникабельность);
- определить языковые понятия высокого уровня, обеспечивающие возможность выполнения однотипных операций.

– Разработанная методика использована для формирования инфологической модели (инфологическая модель - описание предметной области, выполненное с использованием естественного языка, математических выражений, таблиц, графов [3. - С.385]) данных по урожайности и балансу гумуса – важнейших составляющих экономико-математической модели оптимизации структуры сельскохозяйственного. Объект моделирования – СХПК «Новая жизнь» Беловского района Курской области.

Алгоритм расчета урожайности культур на различных категориях пашни построен с использованием ре-

комендованных ВНИИЗПЭ нормативных коэффициентов снижения урожайности относительно планируемого в хозяйстве уровня [2, 4].

В расчете уровней урожайности возделываемых культур на участках пашни различных категорий используется средневзвешенный размерами пашни по агроэкологическим группам коэффициент относительного уровня урожайности данной культуры по сравнению с несмытой пашней (пашней 1-й категории). Для определения урожайности на пашне 1-й категории средняя урожайность по хозяйству делится на вышеуказанный коэффициент, а на пашне других категорий – путем умножения урожайности на немытой пашни на коэффициент относительного уровня урожайности для соответствующей категории пашни (таблица 1).

Таблица 1 – Расчет урожайности сельскохозяйственных культур в СХПК «Новая жизнь» на пашне разных агроэкологических групп*

Культура	Средняя урожайность по хозяйству, ц/га	Средневзвешенный коэффициент относительного уровня урожайности по сравнению с несмытой пашней	Урожайность на пашне 1-й категории (несмытой), ц/га	Коэффициент относительного уровня урожайности на пашне 2-й категории	Урожайность на пашне 2-й категории, ц/га
Пшеница озимая	48	0,980	49	0,90	44
Ячмень	53	0,969	55	0,85	47
Сахарная свекла	516	1,0	516	-	-
Кукуруза на зерно	83	1,0	83	-	-
Подсолнечник	27	0,959	29	0,80	23
Однолетние травы (зеленая масса)	150	0,980	153	0,90	138
Кукуруза на силос и зеленый корм	306	1,0	306	-	-
Многолетние травы (зеленая масса)	180	0,990	182	0,95	173

* Площадь пашни в хозяйстве 4788 га, в том числе пашни 1-ой категории 3810 га, пашни 2-ой категории 978 га.

Таблица 2 - Расчет баланса гумуса при возделывании сельскохозяйственных культур на пашне разных агроэкологических групп в СХПК «Новая жизнь»

Культура	Уравнение для определения массы растительных остатков в зависимости от урожайности, т/га	Баланс гумуса (\pm) по категориям пашни, т/га	
		1	2
Пшеница озимая	$Y = 2,987 + 0,312x$	-0,186	-0,22
Ячмень	$Y = 2,593 + 0,415x$	-0,107	-0,179
Сахарная свекла	$Y = 1,007 + 0,021x$	-1,939	*
Кукуруза на силос и зеленый корм	$Y = 0,65 + 0,107x$	-1,481	*
Кукуруза на зерно	$Y = 0,65 + 0,107x$	-1,844	*
Подсолнечник	$Y = 0,65 + 0,107x$	-1,929	-1,984
Однолетние травы (зеленая масса)	$Y = 3,19 + 0,079x$	0,323	0,315
Многолетние травы (зеленая масса)	$Y = 4,599 + 0,263x$	1,525	1,498
Чистый пар	-	-2,374	-

Масса растительных остатков под различными культурами определяется на основе разработанных во ВНИИЗПЭ статистических функций:

$$y_i = f(x_i),$$

где y_i - количество растительных остатков под i -ой культурой;

x_i - урожайность i -ой культуры.

Накопление гумуса рассчитывается как произведение массы растительных остатков на нормативный коэффициент гумификации. Баланс гумуса для каждой культуры на отдельных категориях пашни определяется как разница между объемом его накопления и количеством минерализованного и смытого гумуса.

Проведенные расчеты показывают, что многолетние травы обладают наивысшей почвозащитной активностью не только в силу антиэррозионной способности,

Условия восстановления почвенного плодородия в упомянутой модели выражаются ограничениями по бездефицитному балансу гумуса на участках пашни различных категорий. Накопление гумуса в почве происходит при гумификации массы растительных остатков возделываемых культур и органических удобрений, а расход гумуса – под воздействием его минерализации и смыва почв.

Коэффициенты положительного либо отрицательного сальдо гумуса на различных категориях пашни по сельскохозяйственным культурам определены по величине урожайности в рассматриваемом хозяйстве (таблица 2).

но и благодаря высокому уровню накопления гумуса в почве. Однолетние травы способствуют накоплению гумуса в значительно меньшей степени, чем многолетние травы. Все зерновые и пропашные культуры и чистый пар формируют в почве отрицательный баланс гумуса.

Таким образом, использование разработанной методики является необходимой составной частью проектирования инфологической модели документальной базы данных для разработки системы компьютерных моделей оптимального планирования сельскохозяйственного производства.

Список использованных источников

- Дейт К. Дж. Введение в системы баз данных / Пер. с англ. 7-е изд. - М.: Вильямс, 2001.

2 Научно обоснованная система ведения агропромышленного производства Курской области / А.П. Щербаков, В.Д. Муха и др. - Курск, 1991.

3 Кузнецов С. Д. Основы современных баз данных // www.citforum.ru.

4 Голицына О.Л., Максимов Н.В, Попов И.И. Базы данных: учебное пособие. - 2 изд., испр. и доп. - М.:ФОРУМ: ИНФРА-М, 2009.- 400 с.

5 Цикритзис Д., Лоховски Ф. Модели данных. - М.: Финансы и статистика, 1985.

6 Методика оптимизации севооборотов и структуры использования пашни. – М.: РАСХН, ВНИЗИЗПЭ, 2004.-77 с.

Информация об авторах

Векленко Василий Иванович, доктор экономических наук, профессор, заведующий кафедрой инновационных методов управления социально-экономическими системами ФГБОУ ВПО «Курская ГСХА», тел. (4712)39-40-15,

Петренко Николай Николаевич, кандидат экономических наук, доцент кафедры инновационных методов управления социально-экономическими системами ФГБОУ ВПО «Курская ГСХА», тел. (4712)53-15-35.

Пашков Владимир Юрьевич, соискатель ФГБОУ ВПО «Курская ГСХА», тел. (4712)53-15-35.

Ерёменко Ольга Викторовна, кандидат экономических наук, доцент кафедры социального страхования, экономики и финансового права Курского института социального образования (КИСО), филиал РГСУ, тел. 8-910-310-00-00.