

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, ОБЧИСЛЮВАЛЬНА ТЕХНІКА ТА АВТОМАТИЗАЦІЯ

УДК 004.725:339.13.025.3

РАЗРАБОТКА АТРИБУТИВНОЙ БАЗЫ ДАННЫХ ПРОСТРАНСТВЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

В.М. Левыкин; Т.В. Гавриш; Е.Е. Малафеев.

Запропоновано підхід до проектування атрибутивного компоненту просторової бази даних на основі методу типологічних угруповань. Розглянуто процедури шкалювання, формування класів в просторі інформативних прикмет-атрибутивів, віднесення об'єктів до класів, які можна використати при розробці логічної інформаційної моделі БД.

Ключові слова: інформація; система; простір; атрибут; шкала; класифікація; ідентифікація; запит.

Разработка современных информационных систем базируется на концепции баз данных (БД). Согласно этой концепции размещаемая в БД информация должна адекватно отображать постоянно изменяющуюся предметную область и удовлетворять информационные потребности потенциальных пользователей. Необходимость выполнения указанных требований должна быть учтена в процессе проектирования информационной системы (ИС).

В рамках широко используемой в настоящее время технологии проектирования SSADM [1] разработке БД предшествует:

- детальное изучение системной среды в терминах предметной области;
- описание внешних объектов, от которых информация поступает в систему и для которых предназначены ее выходные данные;
- изучение внутренних потоков данных, связывающих входы и выходы системы с внешними объектами.

Как правило, независимо от уровня компьютеризации существующей ИС (аналога) указанная информация регламентирована сложившимся в предметной области документооборотом, установленными функциями пользователей, процедурами обработки данных и основными запросами. Последнее существенно облегчает фор-

мирование иерархии потоковых диаграмм и последующий выход на логическую модель данных.

Однако при проектировании получивших в настоящее время достаточно широкое распространение пространственных информационных систем (ПИС) некоторые проектные процедуры и используемый инструментарий должны быть скорректированы. Особенностью ПИС является их работа с объектами и явлениями, для которых обязательна привязка к определенной позиции в пространстве и важны сведения об их пространственной форме, размерах и взаимоположении по отношению к другим объектам. Иными словами, предметная область для ПИС представляет собой совокупность взаимодействующих пространственных объектов. Для описания этой предметной области и последующей разработки пространственной БД необходима, прежде всего, формализация способов задания как отдельных объектов, так и связей между ними. Понятно, что данные способы отличны от используемых в традиционных ИС, хотя общая последовательность технологических этапов проектирования ИС остается прежней.

Спецификой описания пространственных объектов следует считать параллельное использование пространственной и атрибутивной (описательной) информации [2]. Пространственная

інформація включає сведения о положенні об'єкта, задавані з допомогою системи координат або прив'язок, а також топологічні сведения, визначаючі взаємозв'язки в групі об'єктів. Атрибутивна інформація містить як сведения просторового характеру (наприклад, периметр і площа об'єкта), так і характеристики об'єктів, не пов'язані з їх просторовою природою (наприклад, інтенсивність транспортних потоків, тип дорожнього покриття, кількість пересічень магістралей).

Двоїстий характер інформації, описуючої просторові об'єкти, відображається на структурі бази даних. Для порівняно невеликих, але динамічних БД переважно застосовують геореляційний підхід, передбачаючий різну організацію просторового та атрибутивного компонентів БД. Зв'язок між ними підтримується через ідентифікацію об'єкта. Атрибутивна інформація розміщена в таблицях, управління котрими здійснюється звичайними реляційними СУБД. Ці СУБД можуть бути як вбудованими в ПО ПІС як функціональну підсистему, так і бути зовнішніми по відношенню до ПІС. Можлива одночасна реалізація цих двох підходів. В цьому випадку паралельно існують відносно проста вбудована підсистема та зовнішні БД, призначені для зберігання атрибутивної інформації. Таке поєднання найкращим чином відповідає сучасній тенденції пристосовувати ПІС до існуючих під управлінням СУБД баз даних, зміст яких (в розглянутому випадку атрибутивна інформація) модифікується традиційним чином.

Інформаційні потреби користувачів, задоволювані в процесі обробки ресурсів БД, передбачають виконувати системою функції і, відповідно, косвенним чином задають логічну структуру даних (ЛСД). Нагадаємо, що ЛСД є формалізованим зображенням структури інформації, використовуваної в системі, і описує типи зв'язків між об'єктами. Тоді для ідентифікації основних атрибутів об'єктів (своєрив, ознак) необхідно проаналізувати виконувані просторовою інформаційною системою функції. Переважає більшість ПІС реалізує наступні функції [2]:

- оперативні відповіді на порівняно прості (типові) просторові запити;
- підготовка до представлення в потрібному вигляді інформації на більш складні запити для її подальшого використання поза системою (підтримка просторових рішень).

В першому випадку від системи вимагається обслуговування в реальному часі запитів на полу-

чення просторової інформації всього широкого спектра. Це можуть бути заздалегідь сформульовані параметричні запити, що вимагають певного просторового аналізу. Цей режим використовується для підтримки прийняття оперативних рішень, відрізняється чіткою формулюванням і незначительним обсягом вихідної інформації.

Во другому випадку виконуються більш складні просторові запити, що вимагають достатньо серйозної математичної підтримки при обробці вивчаємої з БД інформації (просторово-часова статистика, оверлейний та мережевий аналіз).

Однак незалежно від характеру запитів системою виконуються операції по геокодированню просторових об'єктів (визначення їх місцезнаходження на карті) та обробці тем чи іншим чином інформації, зберігаємої в атрибутивній БД. Крім того, атрибутивні дані, що належать до кожного з просторових об'єктів, повинні містити всю сукупність їх властивостей, необхідних користувачам системи.

Тоді кожна запис (рядок) БД повинна відповідати певному просторово-просторовому об'єкту (наприклад, транспортній магістралі, одиниці адміністративно-територіальної структури міста), а поля (стовпці БД) повинні містити всі атрибути (ознаки) об'єктів, які можуть бути використані при виконанні будь-якого з запитів. Передбачається, що їх перелік та зміст заздалегідь відомі. Кожна запис через відповідний ідентифікатор повинна бути пов'язана з таблицею просторової БД, що задає координати об'єкта.

Виконання запиту зводиться до вибірки з просторової та атрибутивної БД необхідних ознак, проведення з ними певних логічних або вичислювальних операцій та видачі вихідної інформації користувачеві. Остання може бути представлена символьним повідомленням або (і) з допомогою зображення на карті.

В такій постановці завдання побудови ЛСД та реалізуючої її фізичної БД натрапляється на наступні труднощі:

- часто запитувана інформація не може бути представлена ознаками, безпосередньо описуваними просторовими об'єктами;
- кількість ознак, косвенним чином описуваних об'єктами в потрібному аспекті, неоднозначно та залежить від суб'єктивної оцінки особи, що проводить змістовий аналіз предметної області;
- в описанні просторових об'єктів

используются как качественные, так и количественные признаки;

— уровень информативности отдельных признаков, полезных при формировании запрашиваемых данных, неизвестен.

Примером подобной ситуации является запрос на выбор того или иного маршрута при перемещении между точками с заданными координатами. Логично предположить, что предпочтительность маршрута должна определяться характеристиками транспортной инфраструктуры в заданном районе. Однако признака, однозначно описывающего качество транспортной магистрали, не существует. Можно считать, что выбор маршрута должен зависеть от таких факторов, как интенсивность движения, тип дороги, число перекрестков, перепад, высот по ходу движения и пр. Данный перечень свойств не исключает другие варианты по их числу и содержанию. Однако вопрос о вкладе каждого из указанных признаков в процедуру формирования выходных данных должен решаться с учетом их природы и уровня информативности.

С учетом вышеизложенного предлагается для разработки пространственной БД (точнее ее атрибутивного компонента) использовать [3] методы классификации, в частности, типологических группировок, позволяющих из заданной совокупности выделить объекты определенного качества.

Иногда выполнение запроса состоит в поиске в БД соответствующего объекта и отнесение его к одному из заданных классов. Причем, классы (типы) должны соответствовать возможным вариантам ответов на запросы и формироваться с помощью группы информативных признаков, описывающих пространственные объекты в запрашиваемых аспектах.

Иными словами, при проведении классификации пространственных объектов необходимо решить два вопроса:

— выбрать признаки описания объектов в качестве основания классификации;

— сформировать классы и установить правила отнесения объектов к одному из них, т.е. построить кластеры.

При этом решаемую задачу можно интерпретировать следующим образом: необходимо отнести указанную в запросе ситуацию к одному из известных классов, что адекватно нахождению в БД одного или нескольких объектов, в той или иной мере удовлетворяющих запросу.

Следует отметить, что во многих случаях довольно трудно выполнить однозначную классификацию и идентифицировать объекты по одному или даже нескольким признакам. Это связано с тем, что классификация большинства пространственных объектов, имеющих социально-

экономическую или техногенную природу, весьма субъективна, поскольку направлена на достижение определенной цели. В связи с этим во многих практических случаях задачи классификации труднореализуемы и их результативность целиком определяется теми признаками, которые положены в основу.

Для выбора классификационных признаков может быть использован информационно-энтропийный подход [3], основанный на вычислении мер связи между признаками с помощью информационных статистик Шеннона. Однако в большинстве проектных задач выбор параметров базируется на профессиональных знаниях и опыте разработчика системы. В качестве признаков целесообразно использовать, как отмечалось выше, атрибуты пространственных объектов, связанные с содержанием запросов. При этом число признаков непосредственно зависит от четкости задания классов, уровня коррелированности между отдельными признаками, их информативности. При достаточности этих условий число признаков не превышает пяти.

Классификация объектов должна удовлетворять двум положениям: наибольшая однородность объектов внутри класса и возможно большее различие между классами. Естественно, что критерием разбиения объектов на классы является некоторая целевая функция, минимизирующая потери информации при объединении объектов в классы или доставляющая экстремальное значение заданной характеристике объекта (например, минимизация затрат времени и (или) энергетических ресурсов на перемещение между известными пунктами).

В тех случаях, когда формализовать цель и строго описать целевую функцию не удастся, критерием качества классификации служит содержательная интерпретация сформированных групп.

Задача классификации неразрывно связана с измерением признаков объектов, для чего используются различные способы и шкалы. Рассмотрим их особенности.

Предположим, что в результате содержательного анализа всех поступающих в ПИС запросов предложено описывать каждый j -тый, $j = \overline{1, m}$ пространственный объект совокупностью признаков $x(i), i = \overline{1, n}$. Причем, каждое значение $x(i)$ принадлежит некоторому множеству B_i , т.е. $x(i) \in B_i$. Применительно к таблицам БД это означает, что для j -той записи в i -том поле размещается признак $x(i) \in B_i$. Множество B_i содержит l_i возможных градаций признаков

$$x(i) = (x^{(1)}, x^{(2)}, \dots, x^{(l)})_i.$$

Кажда градация характеризует одно из возможных состояний объекта по данному признаку и представляет собой информацию, содержащуюся в i -том поле для j -той записи БД.

Множество B_j может иметь различную природу, характеризующую возможность сравнения и преобразования различных $x(i)$. Соответственно этому следует отметить факт оценки признаков по различным шкалам: номинальной, порядковой и интервальной.

Измерения по простейшей из шкал — номинальной — состоит только в указании градации признака объекта, т.е. отнесении объекта к одному из классов по данному признаку. Номинальные шкалы могут быть построены как для качественных (например, перечень типов дорожных покрытий), так и для количественных признаков (например, номер дороги). Естественно, что числа на такой шкале играют роль идентификаторов и не подлежат арифметическим преобразованиям.

Частным случаем номинальной шкалы является дихотомическая, т.е. шкала, состоящая из двух градаций, обозначающих наличие или отсутствие у объекта признака $x(i)$ (например, факт двустороннего движения).

Если в качестве атрибута используется признак (количественный или качественный), который должен не только задать класс объекта, но и указать в каком отношении он находится с другими объектами, принадлежащими к другим классам, следует остановиться на порядковой шкале. Градации порядковой шкалы упорядочены между собой, как правило, по возрастающей или убывающей степени проявления признака. Причем, оценка по порядковой шкале не содержит информации о величине разности соседних градаций. Это означает, что, если порядковая шкала представлена как $x^{(1)}, x^{(2)}, x^{(3)} \dots$, то информация по каждому из признаков указывает не его числовые значения, а порядковый номер (ранг) объекта по этому признаку. Примером использования порядковой шкалы является их упорядоченность по условиям движения (скорость, рядность).

Для оценки количественных признаков используются интервальные шкалы. Эти шкалы формируются с помощью разбиения области изменения признака на совокупность односвязных непересекающихся интервалов. При этом число интервалов, их величины и расположение фактически задают качество оценки состояния объекта по данному признаку. Определение границ

интервалов, как правило, не формализовано и осуществляется на основе теоретических концепций и опыта прошлых разработок. Таким образом, интервальная шкала позволяет не только указать класс объекта по данному признаку и установить его ранг, но и оценить количественно отличие от других объектов, рассчитав разность (интервал) между соответствующими градациями по шкале.

Формирование интервальных шкал, выполняемое методом естественных интервалов, квантилей, равных интервалов и стандартных отклонений зависит от цели классификации и представляет собой предмет самостоятельных исследований.

Предположим, классификация некоторого пространственного объекта выполняется с помощью n информативных признаков, каждый из которых $x(i)$, $i = \overline{1, n}$ имеет l_i градаций. Тогда раз-

мерность признакового пространства равна $\prod_{i=1}^n l_i$.

Необходимо предложить алгоритм разбиения признаков пространства на K непересекающихся классов. Рассмотрим существующие подходы

Если запрашиваемая информация полностью определяется одним признаком (например, указать наименее загруженную дорогу района в заданном временном интервале), то классификация проводится по этому базовому признаку и от разработчика требуется только обосновать градации используемой шкалы. Последнее определяет достоверность выходного сообщения.

В случае, когда результат запроса зависит от нескольких признаков S , $S \ll n$, то может быть использован [3] алгоритм иерархической классификации. Его суть состоит в последовательном (по градациям каждого признака) разбиении

пространства $\prod_{i=1}^n l_i$ на классы. Причем, последовательность признаков должна быть упорядочена по степени информативности относительно запрашиваемого сообщения.

Вначале образуются классы в соответствии с градациями наиболее информативного признака. Затем каждый из полученных классов разбивается на подклассы градациями следующего по уровню информативности признака и т.д. После последнего этапа разбиения проводится синтез классов, например, путем экспертной оценки выделенных подклассов и объединения некоторых из них в один класс. Наименования полученных классов — это последовательность наименований градаций признаков, расположенных согласно порядку классификации. Например, в класс $\{x_1^{2-4}, x_3^2, x_4^{>3}\}$ входят объекты, имеющие 2, 3 и 4 градации признака 1,

градацию 2 по признаку 3 и градации, превышающие 3, по признаку 4. Признак 2 для поступившего запроса оказался не информативным.

После определения наименования классов не трудно указать, к какому из них принадлежит объект, размещенный в БД. Если один объект попадает в несколько классов, то окончательное решение принимается пользователем с помощью, например, визуализации всех имеющихся место вариантов.

Таким образом, процедуры шкалирования, формирования классов и идентификации в их пространстве объектов фактически задают описание атрибутивных данных, определяют достоверность выходных сообщений и могут быть использованы при проектировании пространственных БД. При этом обработка информации в атрибутивной БД фактически сводится к монотетической классификации объектов по интегральному показателю.

ЛИТЕРАТУРА:

1. Петров Э.Г., Чайников С.А., Овезгильдиев А.О. Методология структурного системного анализа и про-

ектирование крупномасштабных ИУС. — Харьков: Рубикон, 1997. — 140 с.

2. Коновалова Н.В., Капралов Е.Г. Ведение в ГИС. — 2-е изд. — М.: ООО Библион, 1997. — 160 с.

3. Прикладная статистика: классификация и снижение размерности / Под ред. С.А. Айвазяна. — М.: Финансы и статистика, 1989. — 607 с.

УДК 004.725.339.13.025.3

Levikin V.M., a doctor of engineering sciences; Gavrish I.V. a candidate of engineering sciences. The Kharkov university of radioelectronics. Malafeyev E.E., assistant director of joint-stock company AT NIIRI.

Attributive database development of dimensional informational system.

The approach for designing of the attribute component of the dimensional database is offered on the basis of a method of typological classifications. It is possible to utilize surveyed procedures of scaling, creation of classes in space of informative signs - attributes, reference of objects to classes, which at development of a logical entity set model of a DB.

Keywords: *the information; the system, space, attribute, a scale, classification, indefication; inquiry.*

Отримано: 24.02.2003.

Левикін В.М., 2003;

Гавриш Т.В., 2003;

Малафеев Є.Є., 2003.