

Теоретические и методические основы формирования объектно-ориентированной информационной модели навигационной базы данных

Theoretical and methodical bases of formation of an object-oriented information model of the navigation database

Ключевые слова: предметная область – subject area, объектно-ориентированная информационная модель – object-oriented information model, сущность – the essence of the, атрибут – attribute, ER-диаграммы – ER-diagram.

В статье рассматривается методика формирования объектно-ориентированной информационной модели навигационной базы данных.

The article considers the methods of forming the object-oriented information model of the navigation database.

Развитие в геоинформатике объектно-ориентированного подхода к формированию баз данных привело к активным исследованиям в области разработки их информационных моделей. Основное назначение информационной модели заключается в том, чтобы установить взаимосвязь между предметной областью и прикладными задачами (хранения, обработки и т.п.), которые должны решаться на основе этой информации. Поскольку объекты местности могут играть разную роль при решении тех или иных задач или при влиянии тех или иных факторов, то и информационная модель предметной области будет меняться как содержательно, так и структурно. То есть для решения конкретной прикладной задачи необходимо формализовать на уровне концептуального представления (независимо от технического и программного окружения) информационное содержание предметной области [1, 4].

Такое формализованное представление будем называть объектно-ориентированной концептуальной информационной моделью предметной области, которая обеспечивает наиболее удобный интерфейс между прикладными задачами и информацией, хранящейся в базе данных, а также позволяет повысить эффективность использования методов и средств ее обработки в различных

ОСИПОВ / OSIPOV G.

Георгий Константинович

(osipov-g-k-2005@yandex.ru)

доктор географических наук, профессор,
доцент кафедры картографии Военно-космической
академии имени А.Ф. Можайского,
Санкт-Петербург

ПРИСЯЖНЮК / PRISYAZHNIUK A.

Андрей Сергеевич

(pas@itain.spb.ru)

кандидат технических наук,
заместитель генерального директора по экономике
ЗАО «Институт телекоммуникаций»,
Санкт-Петербург

ЕФИМОВ / EFIMOV A.

Алексей Николаевич

старший преподаватель кафедры топогеодезического
обеспечения войск Военно-космической академии
имени А.Ф. Можайского,
Санкт-Петербург

комплексах средств автоматизации. По своей сущности объектно-ориентированная концептуальная информационная модель предметной области должна отражать данные и связи между ними. Если совокупность данных отражает объем и характер обрабатываемой информации, то связи – алгоритм решения этих задач.

На основе вышесказанного под формированием объектно-ориентированной информационной модели базы данных предметной области будем понимать совокупность информации необходимой и достаточной для решения прикладных задач, с использованием создаваемой автоматизированной системы. Следовательно, построение информационной модели должно включать в себя следующие этапы [3]:

- формирование и исследование предметной области;
- идентификация графических и семантических данных и связей между ними;

ГЕОИНФОРМАТИКА

- разработка и исследование локальных объектно-ориентированных информационных моделей предметной области;
- построение обобщенной объектно-ориентированной информационной модели базы данных предметной области.

В основу формирования предметной области должен быть заложен анализ задач, которые предполагается решать с использованием создаваемой автоматизированной системы с целью определения всех элементов данных необходимых и достаточных для их решения. Для этого должна исследоваться каждая задача, путем деления ее на ряд отдельных операций в процессе чего и определяются элементы данных, используемые в них. Это позволяет сформировать матрицу взаимосвязей операций (или задач) и данных, которую можно рассматривать как некоторый формализованный облик предметной области.

Для получения объектно-ориентированной концептуальной информационной модели предметной области необходимо для каждой решаемой задачи производить группировку данных, т.е. определять, какие данные для исследуемой задачи являются сущностью, а какие – атрибутами. Под сущностью будем понимать такой элемент данных предметной области, который раскрывает основное содержание задачи и является в ней ключевым. Атрибут – это свойства сущности. При этом один и тот же элемент данных в одних задачах может выступать как сущность, а в других – как атрибут сущности. Но в пределах одной задачи такие изменения происходить не должны. Это правило должно стать одним из главных при разбиении сложных задач на ряд более мелких подзадач [1].

Для идентификации элементов данных целесообразно использовать метод статистического анализа матрицы взаимосвязей операций и данных, создаваемой в результате исследования предметной области. Этот метод является более эффективным по сравнению с эвристическим, так как при решении задач предметной области используется большое число данных с высокой степенью разнородности и сложными связями. В результате обработки исходной матрицы все данные должны быть распределены на сущности и атрибуты.

Для эффективного решения этой задачи следует менять граничные условия в зависимости от тех правил, которые известны проектировщику базы данных и являются трудноформализуемыми. Это дает возможность получать несколько вариантов идентификации данных, что в конечном итоге позволяет выбирать наиболее

эффективные решения для конкретно решаемой задачи.

Одним из итогов концептуального моделирования является графическое представление модели, которое дает возможность визуально исследовать не только вид и содержание модели, но и ее поведение в случае изменения внешних (изменение задачи, ввод дополнительных данных) и внутренних (изменение граничных условий) факторов. Поэтому графический язык описания концептуальной модели должен быть наглядным, простым и содержательным. Например, конструктивно концептуальную модель можно представить в виде схемы, основными компонентами которой являются сущности, атрибуты и связи. Если сущности описывают основные элементы местности, а атрибуты их свойства, то связи составляют основу знаний о предметной области.

Формализованное описание (аналитическое или графическое) концептуальной модели является объектно-ориентированным.

В случае решения сложных задач представляется целесообразным разбивать их на ряд более простых, что в свою очередь приводит к появлению локальных участков предметной области. Такой подход, с одной стороны, упрощает процесс моделирования, а с другой – позволяет создавать локальные приложения, которые могут независимо использоваться различными задачами. При этом необходимо учитывать, что формирование локальных представлений предметной области требует решения задач поиска компромиссного варианта, так как узкий участок предметной области приводит к снижению уровня интеграции данных и к их «дроблению», а большой участок – к нечеткости и сложности. Исходя из этого, для идентификации данных целесообразно использовать многовариантные решения.

Таким образом, процесс создания обобщенной объектно-ориентированной информационной модели базы данных предметной области сводится к разбиению предметной области на ряд локальных участков, построению для них локальных моделей и объединению локальных моделей путем абстрагирования и обобщения в единую модель предметной области.

Рассмотрим структурную схему построения информационной модели объектно-ориентированной базы данных предметной области [3], рис. 1.

Кратко рассмотрим содержание каждого блока структурной схемы.

Определение содержания объектно-ориентированной базы данных предметной области.

Для определения содержания объектно-ориентированной базы данных предметной



Рис. 1. Структурная схема алгоритма построения информационной модели предметной области

области необходимо решить следующие функциональные задачи:

- определить перечень задач решаемых с использованием разрабатываемой базы данных;
- детализировать каждую решаемую задачу до уровня отдельных операций (подзадач) Z_m ;
- определить совокупность объектов местности, необходимых для решения этих задач;
- определить совокупность данных D_n , которые используются при решении задач;
- сформировать матрицу взаимосвязей задач и данных, которые используются при их решении.

Идентификация элементов содержания объектно-ориентированной базы данных предметной области. Как было сказано выше, для идентификации элементов содержания объектно-ориентированной базы данных предметной области целесообразно использовать метод статической обработки матрицы связи основанный на получении и анализе коэффициентов отражающих связи, частоту использования данных в задачах и их повторяемости.

Для определения степени использования данных в отдельных операциях введем вектор числа использования данных, компоненты которого будем определять путем построчного суммирования матрицы M_{ij} по формуле:

$$K_i^m = \sum_{j=1}^n M_{ij}, \quad (1)$$

где: K_i^m – вектор числа использования данных; j – множество операций, входящих в исследуемую задачу, i – множество данных, используемых при решении этой задачи.

Следовательно, рассчитанный коэффициент характеризует частоту использования данных при решении задачи в целом.

Проведем анализ связей между данными. С этой целью создадим матрицу связи данных с данными – S_{lk} , где l и k равно числу данных, то есть значению параметра i . Исходной информацией для ее создания является матрица M_{ij} . Выделим первый столбец матрицы M_{ij} , который будет соответствовать вектору M_{i1} (взаимосвязь первой операции Z_1 с данными). Выберем такой элемент вектора, для которого выполняется условие $M_{m_11} \neq 0$, где $m_1 \in (1, \dots, i, \dots, I)$. В случае выполнения этого условия примем значение m_1 за номер строки матрицы S_{lk} , то есть $l=m_1$. При этом значение элемента вектора S_{m1k} увеличим на 1 при таком k , когда $M_{k1}=1$. Затем находим следующий элемент m_2 вектора M_{i1} , при котором $M_{m_21}=1$ ($m_2 \in (1, \dots, i, \dots, I)$ и $m_2 \neq m_1$) и выполняем

ГЕОИНФОРМАТИКА

вышеизложенные действия для вектора S_{mjk} . Выполнив эти действия для всех элементов вектора M_{il} , увеличим j на единицу (перейдем к следующей операции). Дальнейшее заполнение матрицы S_{lk} осуществляется по данному алгоритму.

Анализ матрицы S_{lk} позволяет получить вектор полного числа связей между данными по следующей формуле:

$$K_i^S = \sum_{k=1}^l S_{jk} \text{ при } S_{lk} \neq 0 \quad (l = i) \quad (2)$$

Данный вектор показывает степень взаимосвязей между данными, которые используются при решении задач.

Представим значения квадратной матрицы S_{lk} в процентном выражении, где каждый ее элемент будет вычисляться по формуле:

$$S'_{lk} = S_{lk} / K_i^m \times 100\%, \quad (3)$$

Рассчитаем вектор учета использования данных, который определяет число данных в каждой строке матрицы (3), значение которых превышает принятое значение граничного параметра. В качестве, которого выступает параметр «частого использования данных», имеющий значение равное 70 %. Тогда вектор учета использования данных может быть получен по следующей зависимости:

$$K_i^{S'} = \sum_{k=1}^l S'_{lk}, \text{ при } S'_{lk} \geq 70\% \quad (l = i). \quad (4)$$

Для определения вектора коэффициента использования каждого элемента данных K_i^u воспользуемся следующим выражением:

$$K_i^u = K_i^{S'} / K_i^m \times 100\%. \quad (5)$$

В соответствии с полученными значениями векторов K_i^m (1), K_i^S (2) и K_i^u (5) построим гистограммы, основное назначение которых заключается в том, что они должны обеспечить возможность выбрать значения параметров, определяющих частое, среднее и редкое использование данных в задачах, а также совместное использование с другими данными. Кроме того, гистограмма коэффициента использования данных должна дать возможность определить граничные точки, определяющие высокое, среднее и низкое значения коэффициента использования.

Для построения гистограмм выберем значение интервала по оси абсцисс ΔK , который определяется числом связей операций и данных. Исследования показали, что для определения значений интервала можно воспользоваться следующими условиями:

$$\begin{aligned} \text{Для } K_i^m \text{ и } K_i^S & \quad \Delta K = 1, \text{ если } (K_{\max} - K_{\min}) \leq 10, \\ & \quad \Delta K = 2, \text{ если } 10 \leq (K_{\max} - K_{\min}) \leq 20, \end{aligned} \quad (6)$$

$$\Delta K = 5, \text{ если } (K_{\max} - K_{\min}) \geq 20,$$

$$\text{Для } K_i^u \quad \Delta K = 10\%.$$

Анализируя гистограммы определяем граничные точки:

1) для использования данных в операциях:

$$K_{\text{редкое}}^m = 1, \quad K_{\text{частое}}^m = 7;$$

2) для использования данных с другими данными:

$$K_{\text{редкое}}^m = 6, \quad K_{\text{частое}}^m = 7;$$

3) для коэффициента использования данных:

$$K_{\text{редкое}}^m = 6\%, \quad K_{\text{частое}}^m = 61\%.$$

Разрабатаем правила идентификации данных с учетом выявленных граничных параметров:

1. Элемент данных является основной сущностью, если:

– часто используется в операциях, то есть $K_i^m \geq K_{\text{частое}}^m$;

– часто совместно используется с другими элементами данных, то есть $K_i^S \geq K_{\text{частое}}^S$;

– относительно редко используется с другими элементами данных $K_i^u \leq K_{\text{частое}}^u$.

2. Элемент данных является зависимым атрибутом, если:

– не часто используется в операциях, то есть $K_{\text{частое}}^m \geq K_i^m \geq K_{\text{редкое}}^m$;

– не часто используется с другими элементами данных, то есть $K_{\text{частое}}^S \geq K_i^S \geq K_{\text{редкое}}^S$;

– не часто используется совместно с другими элементами данных по сравнению с общим числом его использования во всех операциях, то есть $K_{\text{частое}}^u \geq K_i^u \geq K_{\text{редкое}}^u$.

3. Элемент данных является основным атрибутом, если:

– редко используется в операциях, то есть $K_i^m \leq K_{\text{редкое}}^m$;

– редко используется совместно с другими элементами данных, то есть $K_i^S \leq K_{\text{редкое}}^S$;

– часто используется с другими элементами данных по сравнению с общим числом его использования во всех операциях, то есть $K_i^u \geq K_{\text{частое}}^u$.

Применение этих правил позволяет идентифицировать основные компоненты концептуальной модели объектно-ориентированной базы данных предметной области, то есть определить, какие данные являются основными сущностями, а какие атрибутами зависимой и основной сущности.

Определение зависимых сущностей проводится на основе анализа комбинаций основных сущностей, которые наиболее часто встречаются в ходе решения отдельных операций или задачи в целом.

Оценка соответствия объектно-ориентированной базы данных своему назначению осуществляется путем идентификации связей между операциями и данными, которая выполняется с использованием матрицы связей, где строками являются данные, а столбцами операции.

Для решения этой задачи используется следующий подход. Пусть известны m операций и n данных необходимых для их решения. В процессе анализа базы данных с привлечением экспертов любому объекту, хранящемуся в базе,дается оценка z_{ij} , характеризующая в какой степени i -ый объект обеспечивает решение j -ой операции. На основании полученной информации определяют степень обеспеченности, как каждой операции, так и совокупности операций данными, а также соответствие базы данных в целом своему назначению, для этого используются следующие зависимости:

$$Z_j = \frac{1}{rn} \sum_{i=1}^n z_{ij} \quad (7)$$

$$Z_i = \frac{1}{rm} \sum_{j=1}^m z_{ij} \quad (8)$$

$$Z = \frac{1}{rnm} \sum_{i,j=1}^{mn} z_{ij} = \frac{1}{m} \sum_{j=1}^m Z_j = \frac{1}{n} \sum_{i=1}^n Z_i \quad (9)$$

где: Z_j – степень обеспеченности каждой операции данными; Z_i – степень обеспеченности совокупности операций данными; Z – степень соответствия базы данных в целом своему назначению; z_{ij} – балл, характеризующий в какой степени i -ый объект базы данных, обеспечи-

вает решение j -ой операции; r – максимально возможное значение z_{ij} ; n – число данных, хранящихся в базе; m – число выполняемых операций; j – выполняемая операция; i – объект, хранящийся в базе данных.

При этом базу данных можно считать полностью пригодной для использования по назначению, если для каждого объекта базы будет соблюдено условие:

$$z_{ij} = Z_0 \geq 0 \quad (10)$$

где: Z_0 – некоторое заданное значение оценки (нижний предел пригодности базы к использованию).

Из условия (10) вытекает, что любая из оценок должна быть не ниже заданной величины Z_0 . Идеальным является случай, когда $z_{ij}=Z_{\max}$, т.е., когда любая из оценок получит максимально возможное значение, например, единица, при интервале оценок $\{0;1\}$. Однако такие случаи маловероятны, поскольку объекты базы данных могут вступать в противоречия, которые разрешаются компромиссом, улучшающим одни свойства базы данных за счет ухудшения других. Поэтому необходимо такое преобразование объектов содержащихся в базе данных, которое позволило бы получить оценки z_{ij} , удовлетворяющие условию (10) и максимально приближенные к значению Z_{\max} , т.е. соблюсти условие:

$$\frac{1}{mn} \sum_{i,j=1}^{mn} (Z_{\max} - z_{ij}) = \min \quad (11)$$

В наиболее общем виде оценка соответствия базы данных своему назначению заключается в последовательном выполнении следующих этапов:

- составление программы испытаний и оценочных анкет, подготовка экспертов;
- установление величины Z_0 ;
- испытание и оценка базы данных: определение значений z_{ij} (экспертным методом), вычисление z_i , Z_j и Z ;
- проверка оценок по условию (4), принятие решения по оптимизации содержания базы данных;
- выбор рационального варианта базы данных по наивысшему значению величины Z ;
- выявление наименьших значений z_{ij} , определение слабых сторон содержания, корректировка базы данных;
- повторное испытание и оценка базы данных;
- окончательная доработка базы данных;

ГЕОИНФОРМАТИКА

— определение степени соответствия базы данных своему назначению.

Предложенный подход к оценке соответствия базы данных своему назначению является в достаточной степени объективным и универсальным. Универсальность заключается в возможности его использования для оценки любых баз данных при любом количестве задач и объектов содержания.

Комплексные оценки Z_i и Z_j позволяют выявить слабые стороны тематического содержания и определить целесообразный порядок действий по его улучшению.

Графическое описание концептуальной информационной модели объектно-ориентированной базы данных. Полученная совокупность информационных единиц, определяющих формализованное описание объектно-ориентированной базы данных предметной области, является необходимым и достаточным условием для решения задач анализа и оценки местности. Однако, для установления понятийного интерфейса между предметной областью, комплекса средств автоматизации и его разработчиком необходим соответствующий язык описания концептуальной модели. Для комплекса средств автоматизации этот язык должен быть агрегированный, а для разработчика — графический.

Графический язык представления концептуальной модели должен отображать идентифицированные сущности, атрибуты и связи в терминах графических изображений модели «сущность-связь» [2]. Для описания языка выберем следующие обозначения: основные сущности будем представлять в виде прямоугольника, зависимые сущности — прямоугольника с двойными сторонами, атрибуты — прямоугольником с округленными углами, а связи — ромбами и стрелками. Характер связи обозначается глаголом, вписанным в ромб, а тип связи — количеством стрелок.

Построение ER-диаграммы начинается с отображения локальных представлений. Анализируя сущность, определяют атрибуты, которыми она владеет и связи. После этого сущность ее атрибуты и связи отображают соответствующими обозначениями.

Следующим этапом построения ER-диаграммы является получение схемы, которая описывала бы всю предметную область. Для этого используются методы обобщения, в результате которых происходит объединение локальных диаграмм в одну общую. Такая схема передает информационное содержание обо всей предметной области, а связи являются источником новых

знаний о местности, которая представлена взаимосвязанной совокупностью картографических объектов.

Кроме получения информации о структуре, содержании и знаниях о предметной области, ER-диаграммы позволяют решить проблемы доступа к информации. При этом локальные схемы ориентируют пользователя только к той части информации, к которой он имеет доступ, не зная общей ER-диаграммы, он не может получить доступ к другой информации. Таким образом, общая схема должна быть доступна только администратору, объектно-ориентированной базы данных, а локальные схемы — конкретным пользователям.

Литература

1. Дрождин В.В., Зинченко Р.Е. Системный подход к концептуальному моделированию предметной области в самоорганизующейся информационной системе // Программные продукты и системы. 2009. № 4. С.73–79.
2. Кренке Д. Теория и практика построения баз данных. – 9-е изд. – СПб.: Питер, 2005. – 859 с.
3. Осипов Г.К., Ефимов А.Н. Теоретические основы разработки содержания баз данных о местности в геоинформационных системах военного назначения // Труды военно-космической академии имени А.Ф. Можайского. Выпуск № 631. – СПб.: 2011. С. 130 – 133.
4. Харрингтон Д. Проектирование объектно-ориентированных баз данных. – М.: ДМК Пресс, 2001. – 272 с.