

УДК 004.62:004.492

ПРЕДСТАВЛЕНИЕ ОБЪЕКТОВ ВО МНОЖЕСТВЕ ПРЕДМЕТНЫХ ОБЛАСТЕЙ

Е.В.Малахов

Кандидат технических наук, доцент.
Заведующий кафедрой Информационных
систем в менеджменте Одесского
национального политехнического университета
пр. Шевченко, 1, г. Одесса, Украина, 65044.
Контактный тел.: +38 (0482) 28 - 84 - 17
e-mail: mev@utecon.com

Рассмотрены возможности представления метамodelей предметных областей в виде графов, операций над ними с целью получения новой информации, взаимосвязи и взаимодействия между моделями различных предметных областей

1. Введение

В последнее время значительно возрос интерес к выделяемым предметным областям реального мира как с целью углубления представлений о них, так и с целью расширения многообразия областей. Эти процессы взаимосвязаны друг с другом. Для того чтобы эффективно прогнозировать изменения состояний или развития некоторой предметной области (ПрО), а тем более эффективно управлять ею, необходимо располагать достаточно полной информацией о свойствах сущностей или объектов, «населяющих её», о динамических закономерностях, работающих в конкретной предметной области. Более того, наши априорные представления о предметных областях настолько ограничены, что необходимо накопить достаточно много информации о них, а затем путем анализа и ее обработки увеличить степень полноты, непротиворечивости и определенности.

2. Анализ противоречий в развитии хранилищ данных

Поскольку растущий спрос ведет к увеличению нагрузки на ограниченные ресурсы, успешное использование хранилищ данных требует постоянного контроля приоритетов. Если конкурирующие запросы различных пользователей не могут быть удовлетворены в рамках ограниченного бюджета, то это вызывает разочарование.

Такая проблема может быть вызвана просто текущей ситуацией, связанной с бюджетным циклом, но также может оказаться и индикатором более устойчивого неблагоприятного положения. Хранилище данных должно расти и развиваться вместе с бизнесом. Снижение эффективности вначале может быть очень незначительным и трудно распознаваемым. Для того, чтобы этого избежать, необходимо вовремя увидеть основные признаки такого процесса. Ниже приводятся некоторые из них.

Растущее количество невыполненных заявок на расширение возможностей программного обеспечения. Требования пользователей на предоставление дополнительных данных или запросов не могут быть удовлетворены из-за ограниченных возможностей.

Ограничения на использование хранилища данных. Устанавливаются ограничения на такие параметры, как незапланированные запросы, совмещение операций, окна (периоды) доступа, время выполнения запросов, сложность запросов (отсутствие операций соединения), неприоритетное использование (исследование), дополнительные пользователи и т.д.

Невозможность установки новых приложений. Существуют нужные новые приложения, которые используют уже имеющиеся данные, но платформа не может справиться с дополнительной нагрузкой.

Исторические данные слишком быстро помещаются в архив. Пользователи хотят иметь доступ к более глубоким историческим слоям, но для этого недостает мощности.

Если для успешного нового хранилища данных еще допустимы некоторые ограничения (поступление запросов опережает их выполнение), то для успешного зрелого хранилища непозволительно иметь значительное и непрерывно растущее количество неудовлетворенных требований, поскольку его рентабельность уже установлена и планирование операций более предсказуемо. Если в зрелом хранилище накапливаются невыполненные запросы, то нужно обязательно проанализировать ситуацию, чтобы выяснить причину этого.

Если обнаруживается хотя бы один из вышеназванных признаков, то, возможно, производительность хранилища данных снижается. Причина часто кроется в недостатке масштабируемости в той или иной форме, что, в свою очередь, может быть связано с так называемым *разрывом производительности* (capacity gap) хранилища данных. Он возникает, когда увеличение количества данных опережает возможности платформы хранилища. В результате вначале возрастает стоимость использования хранилища (даже небольшое увеличение производительности стоит все больше), а затем достигается и технологический барьер (т.е. превышаются физические возможности платформы). Обе эти ситуации приводят к тому, что полноценное функционирование хранилища становится невозможным (ограничения использования, задержки выполнения запросов и т.д.) и, таким образом, снижается его производительность.

Использование технологических новшеств – это важная часть общей стратегии, направленной на то, чтобы избежать проблемы снижения производительности Хранилища данных. Вот несколько рекомендаций, как это лучше сделать.

Использовать быстрые улучшения соотношения стоимости/технические характеристики. Сегодня, при быстро появляющихся все новых технологических улучшениях и приспособлении технологий к нуждам потребителей, становится гораздо легче обеспечить хорошее соотношение стоимость/технические характеристики. Проектирование среды Хранилища данных таким образом, чтобы его можно было легко адаптировать к улучшениям соотношения стоимость/технические характеристики, поможет получить лучшую отдачу в рамках имеющегося бюджета на развитие Хранилища.

Снизить количество операций по интеграции. Укрепление открытых стандартов и тенденция развития "предварительной интеграции" означают, что большее количество операций по интеграции может обеспечиваться поставщиками программных средств. Это позволит тратить меньше времени и сил на внедрение новых функциональных инструментов.

Обеспечить масштабируемость платформы хранилища. Недостаток масштабируемости (совмещение операций, сложность, мощность) ограничивает производительность хранилища данных. Выбирая технологию хранилища, нужно быть уверенным, что она может обеспечить линейное увеличение (удвоение ресурсов платформы ведет к удвоению производительности) и расширение (удвоение ресурсов платформы и нагрузки обеспечивает ту же производительность).

Если удастся избежать искушения ограничить использование хранилища данных, а вместо этого применить технологические новшества, обеспечивающие улучшение соотношения стоимость/технические характеристики, то производительность хранилища можно довести до максимума, а его успешное использование будет гарантировано на долгое время.

Одним эффективных способов разрешения указанных противоречий в развитии хранилищ данных является разработка метамодели предметных областей в виде графов.

Однако любая предметная область не существует сама по себе, а окружена и взаимодействует с другими предметными областями, образующими вокруг неё некоторую „среду“. Следовательно, чем больше предметных областей выделено в такой среде и собрано данных о них, тем более полной и адекватной будет информация об анализируемой предметной области.

Как отмечалось в работе [1], для того, чтобы обеспечить полноту описания предметной области (ПрО), очевидно, необходимо определить ее конечные границы. Это можно сделать, например, путём перечисления сущностей, входящих в ПрО, их свойств, функций и связей между ними.

3. Реляционное представление моделей предметных областей

При использовании понятий и подходов реляционной алгебры предлагается каждую сущность представлять многомерной структурой, описывающей его свойства, касающиеся различных аспектов её деятельности или характеризующие её с точки зрения той или иной классификации. Соответственно, описание каждой сущности будет представлять собой трёхмерный куб, состоящий из n плоскостей, где n – максимальное число классификаций конкретной сущности (рис. 1). Причём, подобной структурой описываются все сущности, связанные тем или иным образом с данной. Связь между сущностями может существовать как на уровне всех, так и в пределах некоторых или даже только одной рассматриваемой классификации.

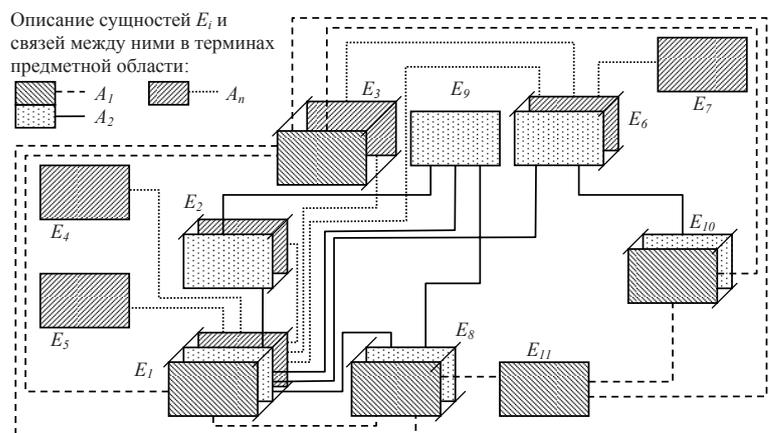


Рисунок 1. Гиперкуб комплексной структуры предметных областей

Таким образом, *предметная область* будет представлять собой множество различным образом связанных трехмерных кубов или *гиперкуб*, описанный некоторым формальным языком L_p , множества сущностей, описанных в терминах одного естественного языка L_N [1].

Если к атрибутам сущности добавить описание его поведения, то получим, говоря языком объектно-ориентированного анализа, *объект* некоторой предметной области.

Таким образом, информационное описание предметной области A_m сводится к определению множества объектов этой предметной области $E(m) = \{E_1(m), \dots, E_i(m), \dots, E_n(m)\}$. Причем, каждый такой объект описывается, множеством свойств $X_i(m) = \{x_{i1}(m), \dots, x_{ij}(m), \dots, x_{ik}(m)\}$, которое при представлении в реляционной базе данных соответствует схеме отношения, отражающей этот объект и множеством зависимостей от других объектов или зависимостей, определяющих другие объекты, т.е. связей $R(m) = \{R_{ij}(m)\}$.

Следовательно, $A_m = \{E(m), R(m)\}$, что практически соответствует выражению, используемому в классическом определении графа. Поэтому очевидно, что представление метамодели предметной области возможно с использованием не только теории реляционных баз данных, но и теории графов.

4. Графовое представление моделей предметных областей

При использовании теории графов каждая предметная область представляется в виде графа (рис. 2 и рис. 3), вершинами которого являются объекты (точнее их описания или метайнформация, которая на рис. 1 была представлена таблицами), а ребрами — связи, функциональные отношения и т.п. между ними. В графовом представлении метамодель будет выглядеть так, как показано на рис. 2. При таком представлении видно, что каждая плоскость (реляционная таблица), описывающая соответствующий объект реального мира (рис. 1), представляет собой проекцию этого объекта $E_i(j)$ на соответствующую предметную область A_j (рис. 2): $E_i(j) \in A_j$.

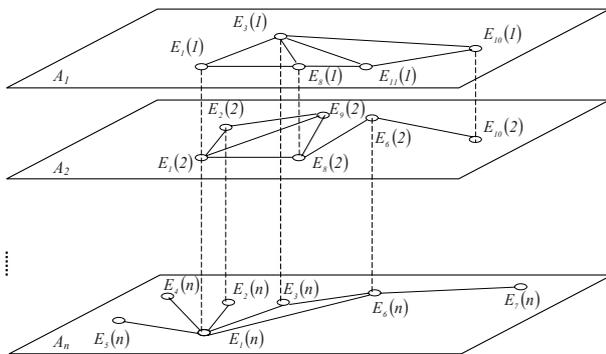


Рисунок 2. Графовое представление комплексной модели предметных областей

При таком описании можно говорить об n -мерном информационном представлении реального мира или какой-либо его части, где n — это всё множество предметных областей, на которые спроецировано всё множество известных объектов. Причем, если проекция E_k на $PrO A_r$ отсутствует, то для целостности представления будем считать, что $E_k(r) = \emptyset$. Соответственно, для примера на рис. 2:

$$E_2(1) = E_3(2) = E_4(1) = E_4(2) = E_5(1) = E_5(2) = E_6(1) = E_7(1) = E_7(2) = E_8(n) = E_9(1) = E_9(n) = E_{10}(n) = E_{11}(2) = E_{11}(n) = \emptyset$$

Аналогично понятию подграфа [2, С. 25], некоторое подмножество проекций связанных объектов на какую-либо $PrO A_j$, представленных смежными вершинами соответствующего графа, будем называть i -й подобластью (ППрО) j -й PrO : $S_i(j) \subseteq A_j$.

Используемый для описания модели PrO математический аппарат определяет набор операций, которые можно выполнять над метаданными.

Очевидно, что при реляционном описании метаданные представляются отношениями. Соответственно, операциями над ними будут операции реляционной алгебры, модифицированные с учётом того, что операндами будут метайнформация. В свою очередь, такой подход позволит использовать существующие СУБД и их функции и команды манипулирования таблицами и хранилищами для выполнения операций над метаданными.

В случае использования теории графов операциями над метаданными будут некоторые операции над

графами, адаптированные с учетом того, что операндами является метайнформация об объектах предметной области [2].

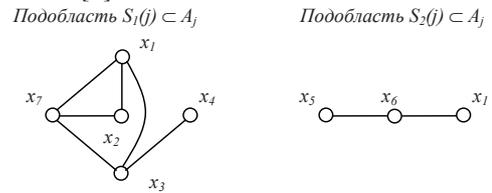


Рисунок 3. Подобласти S_1 и S_2 предметной области A_j

Выполнение таких математических операций над графами, описывающими предметные области (или подобласти некой предметной области), позволяют:

1. Создать расширенную (полную) модель PrO (например, операцией объединения, рис. 4) или получить многомерную структуру, подобную представленной на рис. 2, что, в свою очередь, составит более целостную картину реального мира и/или даст возможность „увидеть” объекты, проекции которых, возможно, были „потеряны” в других плоскостях.

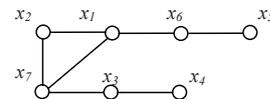


Рисунок 4. Полная модель предметной области A_j

2. Выделить или синтезировать объекты (и связи между ними), которые являются характерными представителями PrO или некоторой подобласти данной PrO (операции пересечения и дополнения).

Множество объектов, полученных в этом случае, составит, т.н., объектное ядро PrO (в примере рис. 3 это будет только один объект x_7). Кроме того, такие операции дадут, очевидно, те объекты, информация о которых может быть вынесена в информационное хранилище того или иного уровня.

3. Синтезировать новую (виртуальную) предметную область, „населённую” только теми объектами, которые интересуют исследователя.

Что касается „потерянных” проекций, то например, можно предположить, что вершины (и ребра), которые приводят к нарушению изоморфизма подграфов, принадлежащих различным плоскостям $S_i(j)$ и $S_i(k)$, представляют собой „недостающие” проекции объектов (и их связей) на соответствующие PrO . Причём вероятность наличия такой „потери” возрастает пропорционально числу m плоскостей, в которых представлен подграф S_j , и степени r пропущенной в k -м подграфе вершины $E_i(k)$ в тех подграфах, где она представлена:

Подобласть $S_i(j)$

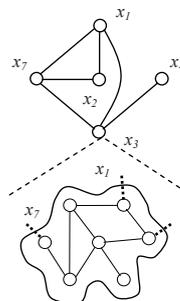


Рисунок 5. Детализация объекта предметной области A_j

$$P(E_i(k) \neq \emptyset) = 1 - \frac{1}{m \cdot r}$$

Если предположить, что каждый объект PrO состоит из множества элементарных объектов (рис. 5) [4, 5], то в результате операции над подграфами такой детализированной модели мож-

но виділити елементарні об'єкти, формуючі об'єктний базис відповідної ПрО або її підобласті.

Виділення такого ядра, в свою чергу, дозволить синтезувати нові об'єкти, які, по крайній мірі, теоретично будуть мати право на існування в даній ПрО.

Структурою, аналогічною наведеної на рис. 5, можна представити і метаінформацію, описуючу кожен об'єкт, або, як мінімум, його властивості. Виконуючи відповідні операції можна, також як і з допомогою елементарних об'єктів, синтезувати нові або інтегровані об'єкти.

5. Висновки

Представлений спосіб попередньої селективної обробки інформації для створення метамодель предметних областей в вигляді графів дозволить суттєво підвищити ефективність формування і використання сховищ даних. Разом з тим будь-яка предметна область не існує сама по собі, а оточена і взаємодіє з іншими предметними областями, оточуючими навколо неї певну „середу”. Тому, чим більше предметних областей виділено в такій середі і зібрано даних про них, тим більш повною і адекватною буде інформація про аналізовану предметну область.

Література

1. Востров Г.Н. Проблеми опису структури предметних областей / Г.Н.Востров, Е.В.Малахов, В.Н.Кулешов // Тр. Одес. політехн. ун-та. – Одеса: ОНПУ, 2000. – Вип.2 (11). – С.111–114.
2. Татт У. Теорія графів: Пер. з англ. – М.: Мир, 1988. – 424 с.
3. Малахов Е.В. Проблеми використання метаданих в цілях перетворення предметних областей / Малахов Е.В., Добровольська Т.В.// Холодильна техніка і технології. – Одеса: ОДАХ, 2005. – Вип.6 (98). – С. 112-116.
4. Малахов Е.В. Оцінка ступеня адекватності баз даних як інформаційних моделей предметних областей // Тр. Одес. політехн. ун-та. – Одеса: ОНПУ, 2004. – Вип. 1 (21). – С.82-86.
5. Востров Г.Н. Базис простору об'єктів предметних областей при проектуванні бази даних / Востров Г.Н., Малахов Е.В.// Тр. Одес. політехн. ун-та. – Одеса: ОНПУ, 2005. – Вип.1 (23). – С. 78 – 81.

УДК 002

АНАЛІЗ СУЧАСНИХ МЕТОДОЛОГІЙ СТВОРЕННЯ ІНФОРМАЦІЙНИХ СИСТЕМ, ПРИЗНАЧЕНИХ ДЛЯ ОБЛІКУ ТА АНАЛІЗУ ІНФОРМАЦІЇ

С.Л. Пелевін

Аспірант заочної форми навчання Київського
Національного університету будівництва та архітектури
Контактний тел.: 8 (067) 321-36-62
e-mail: s_pelevin@am-soft.com.ua

Г.А. Гайна

Кандидат технічних наук, доцент
Київського Національного університету будівництва та
архітектури

Проведена класифікація та аналіз сучасних методологій, придатних для створення систем обліку та аналізу інформації, розглянуті проблеми рівня формалізації методологій та підходи розробки, що використовуються у методологіях. Сформовані критерії вибору методологій виходячи з характеристик програмного забезпечення, сфери застосування тощо

Вступ

Тенденції розвитку сучасних інформаційних технологій приводять до постійного зростання складно-

сті інформаційних систем, що створюються у різних галузях. Сучасні великі інформаційні системи обліку та аналізу інформації характеризуються, як правило, наступними особливостями: