УДК 004.655

В.В. Осипова, И.Л. Чудинов

Проектирование концептуальной информационной модели предметной области на основе анализа доменов атрибутов описания объектов в рамках реляционной модели данных

Предложен формализованный подход для проектирования концептуальной информационной модели предметной области на основе анализа доменов атрибутов описания объектов в рамках реляционной модели данных с учетом интеграции информационных потребностей пользователей.

Ключевые слова: проектирование информационной модели, интеграционный подход, анализ доменов атрибутов отношений.

В настоящее время качество современных информационных систем (ИС) в значительной степени определяется результатом проектирования базы данных (БД), являющейся концептуальной информационной моделью предметной области (КИМПО). В процессе проектирования БД анализ предметной области (ПрО) и определение ее КИМПО является самым сложным и трудно формализуемым этапом, поэтому любые подходы, направленные на его формализацию, весьма актуальны [1].

Традиционные подходы опираются на инфологическое (вербальное) представление ПрО, которое служит основой для интуитивно понятного формирования объектов ПрО – сущностей. При этом предполагается тесное взаимодействие специалистов в области проектирования ИС и квалифицированных сотрудников ПрО. Это может быть приемлемым для относительно простых ПрО, таких как часто используемые в методиках проектирования описания деятельности фирм по туристическому бизнесу, продаж билетов на железнодорожные или авиамаршруты и подобных узко специализированных сфер деятельности [2]. Для сложных ПрО таким способом можно выявить лишь независимо составленные описания деятельности отдельных функциональных подразделений, причем число и детальность таких описаний расширяются с течением времени. Применение такого подхода к проектированию КИМПО как совокупности локальных моделей приводит к созданию автономных ИС для функциональных подразделений («Абитуриент», «Кадры», «Деканат», «Аспирантура» и т.п.) и, как следствие, к серьезным проблемам, связанным с дублированием процедур сбора и хранения первичных данных. Такое положение было в какой-то степени вынужденным в 80-е годы в условиях отсутствия корпоративных вычислительных сетей. Сегодня эта проблема в значительной степени преодолена, и создается возможность использования безызбыточной единой информационной модели [3]. Однако сложность такой модели не позволяет исключить автономного проектирования фрагментов КИМПО. Предлагаемый подход не только предполагает аналогичное автономное проектирование конкретных сущностей, но и как бы усиливает эту автономность в части определения домена каждого атрибута для определения наличия и типа связей между сущностями ПрО. Предлагаемый подход рассматривается в рамках реляционной модели данных (далее «отношение» соответствует «сущности»), являющейся в настоящее время основой абсолютного большинства ИС [4]. Таким образом, преимуществом применения этого подхода является высокая степень его формализации, что позволяет автоматизировать процесс проектирования практически любой ПрО. Для формирования КИМПО систем организационного управления, которые, как правило, являются большими и сложными ПрО, к числу которых можно отнести и вузы, необходим анализ существующих информационных потребностей пользователей, на котором и основан предлагаемый интеграционный подход, базирующийся на следующих принципах.

Связь между отношениями существует, если может быть выполнена операция соединения, реализация которой, в свою очередь, возможна на основе сопоставления значений атрибутов соединяемых отношений (назовем их атрибутами связи). Причем определяющее значение имеет совпадение значений атрибутов связи, что в свою очередь может интерпретировать связь между отношениями как отнесение их к одному классу, а признаком класса – значение домена атрибута.

Полагается, что связи между парами отношений в концептуальной модели соответствуют общепринятым 1:1, 1:M, M:M, причем связь M:M преобразуется к двум связям 1:M. Тогда, с учетом первого принципа атрибутами связи на стороне 1, должны быть

атрибуты полного ключа, а на стороне М – либо атрибуты части ключа, либо неключевой атрибут. Преобладание связей 1:М в КИМПО если и не обосновывается, то, по крайней мере, хорошо согласуется с иерархическим подходом и людей к анализу и синтезу (построению) различных объектов материального и виртуального мира, а значит, и к построению информационных моделей ПрО.

Автономно (независимо друг от друга) сформированные отношения могут интегрироваться в КИМПО (с учетом второго принципа) прежде всего на основе анализа доменов ключевых атрибутов, что позволяет не только определить тип связи 1:1, 1:М, М:М, но и некоторые свойства связи (идентифицируемость, необязательность).

Иерархическая связь 1:М наследуется, т.е. если между отношениями A и B существует связь типа 1:М и между отношениями B и С также имеется связь типа 1:М, то и между отношениями A и C следует наличие связи 1:М. Это позволяет сократить число анализируемых пар отношений и построить оптимальную иерархическую модель.

Формализованная основа предлагаемого подхода заключается в следующем. Для определения связей между отношениями используется анализ отношений доменов атрибутов этих связей. Анализируя домены $D1$, $D2$ атрибутов $A1$, $A2$ отношений $R1$, $R2$, соответственно, определяем, являются ли атрибуты $A1$ и $A2$:

– совместимы, если объединение доменов ($D1 \cup D2$) имеет смысл в качестве домена в исследуемой ПрО;

– сопоставимы, если их домены пересекаются ($D1 \cap D2 \neq 0$);

– соподчинены, если домен одного атрибута входит в состав другого ($D1 \subset D2$ или $D2 \subset D1$);

– совпадают, если их домены полностью совпадают по составу ($D1 = D2$).

Далее, на основе введенных определений о сопоставимости атрибутов, задаем условия связуемости отношений. Полагаем, что отношения $R1$ и $R2$ являются связуемыми, если они имеют хотя бы одну пару соответствующих (по крайней мере, сопоставимых) атрибутов. Учитывая, что основным средством отображения связи между отношениями являются ключи, распространим условие связуемости на ключи $K1$, $K2$ отношений $R1$, $R2$ соответственно. Тогда ключи $K1$ и $K2$ связуемы, если они имеют хотя бы одну пару соответствующих (по крайней мере, сопоставимых) атрибутов. Назовем ключи $K1$ и $K2$ равноценными ($K1 = K2$), если эти ключи имеют один и тот же ранг ($g = p$) и все их атрибуты попарно соответствуют друг другу. Если ключи имеют разные ранги ($g < p$), то считаем ключ $K2$ зависимым от ключа $K1$ ($K2 < K1$), т.е. ключ $K2$ имеет подмножество $K2' \subset K2$, равноценное ключу $K1$ ($K1 = K2'$). Ключи $K1$ и $K2$ являются пересекающимися ($K1 \times K2$), когда они имеют подмножества попарно соответствующих атрибутов ($K1' \subset K1$, $K2' \subset K2$ и $K1' = K2'$).

Учитывая непрерывное развитие КИМПО, связанное с появлением новых информационных потребностей, необходима их корректная (с точки зрения исключения избыточности) интеграция с существующей. Идея алгоритма интеграции новых информационных потребностей с существующей КИМПО основывается на том, что отношения текущей информационной модели сгруппированы в множества с одинаковыми количествами атрибутов в ключе, вновь созданные отношения интегрируются в текущую модель по следующему алгоритму. Алгоритм построен таким образом, что некоторое отношение (анализируемое) последовательно сравнивается с множеством других уже включенных в базовую КИМПО отношений (исследуемых).

Далее связываем отношения внутри каждого класса, начиная с первого. Последовательно проверяется связуемость отношений с уже исследованными в своем классе.

Если ключи анализируемого и исследуемого отношений совпадают, то производится соединение анализируемого исследуемым путем объединения их схем.

Если ключи равноценны, но не совпадают, то между ними устанавливается связь типа соединение.

Если ключи анализируемого и исследуемого пересекаются, то выполняется предыдущий этап в классе, соответствующем ключу пересечения. Если и там не найдено равноценных ключей, но есть пересечение, то аналогично осуществляется «подъем» в класс, соответствующий ключу пересечения.

Если на каком-то уровне будут обнаружены отношение с равноценным ключом, то с ним, как со старшим, устанавливается иерархическая связь исходных отношений как подчиненных (произошло преобразование связи типа группировка).

Если ни на одном из уровней не будет обнаружено связи ни с одним отношением, то в состав исследуемых включается виртуальное отношение, состоящее из ключа, совпадающего с пересечением ключей исходных отношений. Для аналитика это сигнал к поиску в ПрО объекта описания с таким ключом.

Аналогичным образом для каждого отношения i-го класса ($i \geq 2$) проводим анализ на связуемость отношений с отношениями $i+k$ класса.

Следует отметить, что полученные связи отношений могут обладать свойством наследования: если отношения $R1$ (старшее) и $R2$ имеют такой же тип связи, что и связь между отношениями $R2$ (старшее) и $R3$, то $R3$ наследует тип связи $R1$. Тогда для наследуемых связей не имеет смысла применять описанный алгоритм.

Таким образом, на основе интеграции информационных потребностей пользователей предложенный подход позволяет сформировать КИМПО, применяя формальный аппарат связуемости отношений (сущностей), используя домены атрибутов в качестве фундаментальной основы для формализованного вывода отношений между сущностями.

Литература

1. Navathe S.B. Evolution of Data Modeling for Databases // Communications of the ACM. – 1992. – Vol. 35, Issue 9. – P. 112–123.

2. Диго С.М. Проектирование и использование баз данных. – М.: Финансы и статистика, 1995. – 208 с.

3. Чудинов И.Л. К вопросу о проектировании концептуальной модели систем организационного типа // Кибернетика и вуз. – Томск: Изд-во ТПУ, 1999. – Вып. 29. – С. 10–24.

4. Godd E.F. A Relation Model of Data for Large Shared Data Banks // Communications of the ACM. – 1970. – Vol. 13, Issue 6. – P. 377–387.

Осипова Виктория Викторовна
Аспирант каф. оптимизации систем управления
Национального исследовательского Томского политехнического университета (НИТПУ)
Тел.: (382-2) 56-33-94
Эл. почта: vikosi@tpu.ru

Чудинов Игорь Леонидович
Канд. техн. наук доцент каф. оптимизации систем управления НИТПУ
Тел.: (382-2) 56-33-94
Эл. почта: chil@tpu.ru

Osipova V.V., Chudinov I.L.

Designing a conceptual information model of knowledge domain using the analysis of the object attributes' domains in terms of relational data model

The formalized approach for designing a conceptual information model of knowledge domain, which using the analysis of the object attributes' domains in terms of the relational data model taking into account the integration of users information needs, is suggested.

Keywords: information model designing, integration approach, analysis of relation attributes' domains.