

УДК 519.85

МЕТОД ПОСТРОЕНИЯ МАТЕМАТИЧЕСКОЙ МОДЕЛИ ПРЕДМЕТНОЙ ОБЛАСТИ

© 2010 В.В. Антонов, Г.Г. Куликов*

Ключевые слова: атрибутивная трансляция, предметная область, семантическая модель, методология структурного анализа.

Выявлены проблемы построения математической модели предметной области. Эволюцию информационных систем можно рассматривать с позиции их влияния на организацию: встраивание информационной системы в действующую бизнес-форму или ее изменение. Доля затрат на получение и переработку информации в экономике и общественных расходах непрерывно возрастает. Однако остаются неисследованными многие теоретические аспекты, связанные с моделированием процессов автоматизации и их дальнейшим информационным сопровождением совместно с семантическими правилами, их регламентирующими. Предложен вариант моделирования предметной области с позиций методов, учитывающих нечеткость или неопределенность описаний модели исследуемого объекта.

Мы живем на рубеже перехода к информационной эпохе, перехода к внедрению эффективных методов управления сложными и быстропротекающими процессами, изменения методов управления не только технологическими процессами, но и хозяйственной и социальной деятельностью, что стало возможным благодаря бурному развитию информационных технологий. В результате повсеместной информатизации многие функции управления передаются под контроль сложноорганизованных информационных систем, использующих биологические и компьютерные технологии обработки информации. Обработка информации в подобных системах стала самостоятельным научно-техническим направлением. Для данных систем характерно наличие технологических участков с автоматическим, автоматизированным и интеллектуальным управлением. Построение информационной системы, переход из одного состояния в другое происходит по определенным закономерностям, при этом закономерности переходов не всегда могут задаваться в четком виде, т.е. система обладает поведением, определяемым закономерностями. Закономерность целостности проявляется при каждом состоянии системы, благодаря чему при каждом состоянии системы могут возникать новые свойства, которые не могут быть выведены как сумма свойств элементов.

Перевод условий практической задачи на язык математических моделей всегда был трудным и зачастую приводил к потере трудно-формализуемой качественной информации. Многие современные задачи управления просто не могут быть решены классическими методами из-за очень большой сложности математических моделей, их описывающих. В условиях применения автоматизированных систем происходит трансформация функций человека, возникают новые связи между человеком и системой, некоторые функции полностью передаются системе. Компьютеризация способствует расширению возможностей субъекта, порою качественно меняя содержание его деятельности. И эта деятельность должна быть четко определена и закреплена за субъектом так же, как и за автоматизированной системой. Модель, охватывающая информационную систему, может быть представлена в виде метабазы, в которой содержится информация по каждому виду объекта учета. С другой стороны, информационная система представима в виде функциональной системы, т.е. в виде множества функций. Таким образом, цели и ограничения могут быть заданы как нечеткие множества. Взаимосвязь между ними может быть определена отношением на декартовом произведении¹. Рассматривая цели и ограничения в виде симметричных элементов логи-

* Антонов Вячеслав Викторович, кандидат технических наук, доцент Уфимского государственного авиационного технического университета. E-mail: boss@bashkortostan.ru; Куликов Геннадий Григорьевич, доктор технических наук, профессор, зав. кафедрой АСУ Уфимского государственного авиационного технического университета. E-mail: gennadyg_98@yahoo.com.

ческой схемы, можем достаточно просто сформировать на их основе решение, которое является по существу выбором одной или нескольких из имеющихся альтернатив. При этом нечеткое решение может рассматриваться как некоторая “инструкция”, нечеткость которой является следствием неточности формулировки поставленных целей и ограничений, т.е. влияние нечеткой цели и нечеткого ограничения на выбор альтернатив характеризуется их пересечением, которое образует нечеткое множество решений.

Бизнес-процессы, рассматриваемые как функциональная модель реальных процессов, представляют собой структурированное описание заданной последовательности выполняемых бизнес-операций, т.е. горизонтальную иерархию внутренних и зависимых между собой функциональных действий. При этом бизнес-процесс, рассматриваемый как совокупность последовательно выполняемых цепочек операций, может трактоваться как множество взаимодействующих подсистем, т.е. в виде дискретной динамической системы изменяемой в пространстве и времени. Моделирование информационной системы имеет общую философскую основу. Наиболее существенным для моделирования философским понятием является предметная область, которая может быть определена как мысленно ограниченная область реальной действительности или область идеальных представлений, подлежащая моделированию, состоящая из объектов, находящихся в определенных отношениях между собой и имеющих разные свойства. При этом под свойством понимается характерная особенность объекта, для оценки которой устанавливается определенная мера - показатель свойства, который характеризуется множеством значений. Эти значения обозначаются символами из некоторого заранее определенного множества, называемого алфавитом. Таким образом, свойство объекта является реальностью, а показатель - субъективной мерой этой реальности. Совокупность показателей свойств объекта является параметрами объекта. Общепризнанное формальное определение понятия предметной области до настоящего времени отсутствует. Определение предметной области как части реального мира или совокупности классов реальных объектов

подлежащих моделированию, предполагает модельное отражение с целью изучения под определенным углом зрения. Этот угол зрения сам входит в понятие предметной области. Поэтому большинством исследователей принято считать, что понятие предметной области не может быть формализовано как первичное понятие. В последние годы в теории баз данных и информационных хранилищ сформировались новые направления научных исследований, получившие название теории концептуального моделирования в базах данных, корпоративных хранилищах данных. Последующим развитием создаваемой теории стало развитие методов моделирования предметных областей в информационных системах. Модель предметной области - это наши знания о предметной области. При исследовании предметной области может быть получено значительное количество информации, которая носит субъективный характер. Ее представление на естественном языке содержит нечеткости или неопределенности, которые не имеют аналогов на языке традиционной математики. При определении понятия предметной области необходимо учитывать следующий методологический аспект, при рассмотрении предметной области как части действительности - онтологический, а при рассмотрении предметной области как знания про эту действительность - гносеологический. И как результат имеем два различных класса моделей и задачу поиска соответствия данных моделей действительности - модель действительности и модель знаний про эту действительность.

Основная проблема состоит в неформализованности процесса моделирования предметной области, что делает невозможным применение математических методов анализа свойств моделей предметной области, таких как функциональная полнота и целостность. Все это ставит вопрос о рассмотрении задачи моделирования предметной области с позиций методов, учитывающих нечеткость или неопределенность описаний модели исследуемого объекта. Говорить о предметной области имеет смысл, если она имеет определенную семантическую локализацию, например в пространстве и времени или функциональную. Тогда построение семантической модели сводится к формализации логи-

ческих отношений. Существуют разнообразны́е методы принятия решений, представляющие различное упорядочение рассматриваемых решений основанных на одних и тех же экспертных оценках. Одна из основных сложностей при разработке модели предметной области состоит в том, что число возможных вариантов формализации предметной области бесконечно. Модель должна адекватно отображать любое подмножество (вариант формализации предметной области), а сам процесс моделирования может иметь любую идею, позволяющую определить значение любого объекта предметной области путем реализации какой-либо определенной последовательности действий. Одна из таких идей - метод синтаксически-ориентированной трансляции, основанный на работах Ноэля Хомского. Из гипотезы Хомского² следует, что семантический анализ может быть сведен к синтаксическому и состоит из двух шагов - распознавания структуры и построения выходных действий на основе этой структуры. Таким образом, математический подход позволяет ограничиться множеством цепочек, которые можно определить некоторым точным образом. То есть можем говорить о некотором формальном языке, заданном в виде множества. Для построения необходимо иметь алгоритм, который по заданной грамматике строит вывод, порождаемый этой грамматикой. Согласно такого алгоритма для произвольной грамматики Хомского не существует³. Возможно несколько путей решения данной проблемы, заключающихся либо в разработке алгоритма распознавания для каждого частного случая, либо наложения ограничений на правила грамматики, выделения подклассов грамматики, для которых алгоритм существует. Выделяя объекты в предметной области, получаем, что каждый объект определяется конечным множеством атрибутов, каждый атрибут имеет множество допустимых значений, и мы имеем атрибутивно транслируемую грамматику, согласно которой определяются правила вычисления атрибутов и может быть построен алгоритм атрибутивного дерева вывода. Д. Кнут формализовал подобные идеи, введя понятие "атрибутивной трансляции"⁴. Символам грамматики ставится в соответствие конечное число семантически определенных атрибутов, зна-

чения которых могут быть числовыми, символическими, матрицами или какими либо признаками. Правила грамматики определяют связи этих атрибутов, а вычисление значений определяется деревом вывода. Все это позволяет определить семантику языка так чтобы значение любого атрибута, в любом узле дерева вывода могло произвольным образом зависеть от всего дерева. Отсюда следует, что любая функция, которая определена на дереве вывода, может быть представлена в виде атрибута какого-либо определенного узла⁵. Предметную область можно декомпозировать на элементарные объекты, каждый из которых описывается совокупностью атрибутов. Объекты предметной области связаны между собой определенными отношениями, которые можно в совокупности представить в виде взвешенного по ребрам частично ориентированного графа. Структура графа представляет структуру предметной области. Подграфы графа представляют сложные объекты или подсистемы предметной области. Вместо графов для представления структуры предметной области можно использовать язык теории множеств и решеток их разбиений. Каждый кортеж базы данных является описанием состояния некоторого элементарного объекта. Подмножество всех кортежей, сходных с данным кортежем относительно выбранной меры сходства, является представлением элементарного объекта.

Пусть некоторый объект выделен и обозначен термином. Как правило, новый объект сравнивается с уже известными объектами, и его информационная модель формируется как совокупность сравнения информационных моделей ранее известных объектов. При этом модель новой предметной области для данного объекта будет строиться на основе предметной области того объекта, который стал известным первым. В результате знания о предметной области, включающей в себя данный объект, будут структурированы в виде совокупности свойств первого выделенного объекта и последовательности изменений предметных областей последующих объектов. Пусть W - исследуемая предметная область объектов $W = \{w_1, \dots, w_n\}$. Часто в моделях рассуждений присутствуют нечеткие понятия, однако любая информация, записанная в ка-

ком-либо формализованном виде и представленная в памяти компьютера является четкой. Поэтому нечеткость знаний или отношений может определяться семантикой самой информации. Не включенные в выделенные предметные области свойства можно рассмотреть как отдельную предметную область с особыми свойствами - окружающую среду. Таким образом, любая предметная область может считаться распределенной, так как она взаимодействует с окружающей средой. Согласно вышесказанному, $w_{i+1} = F_i^W(w_i)$, где

$F_i^W(w_i)$ не является функций в обычном понимании, а определяет лишь возможное состояние предметной области одного объекта на основе отличия от другого. То есть в данном случае речь идет о возможно нечетких связях между метаобъектами.

Модель предметной области определяется посредством функции представления и семейства моделирующих функций. Пусть S - модель предметной области объектов

$F^M : W \rightarrow S$. Каждому объекту предметной области может быть поставлена в соответствие модель (на основании функции моделирования F^M) $r_i = F_i^M(w_i)$. Таким образом, каждому объекту $w_i \in W$ соответствует $r_i \in S$.

При этом должна присутствовать функция $F_i^R(r_i)$, которая однозначно определяет r_{i+1}

по r_i , т.е. $r_{i+1} = F_i^R(r_i)$. Очевидно, что r_{i+1}

может быть определено по цепочке $w_{i+1} = F_i^W(w_i)$ и $r_{i+1} = F_{i+1}^M(w_{i+1})$. После произведенной подстановки получаем

$r_{i+1} = F_{i+1}^M(F_i^W(w_i))$. С другой стороны, существует, по меньшей мере, еще одна цепочка

$r_i = F_i^M(w_i)$ и $r_{i+1} = F_i^R(r_i)$, согласно которой после проведенных подстановок получаем

$r_{i+1} = F_i^R(F_i^M(w_i))$. Сопоставляя полученные результаты, можем сделать вывод, что независимо от того, выполнена ли сначала операция в предметной области, а затем произведено отображение на модель предметной области, или сначала произведено отображение на модель, а потом в модели предмет-

ной области выполнена соответствующая операция, результат будет одинаков. Следовательно, для каждого зафиксированного значения i получаем гомоморфизм F_i^W и F_i^R .

Таблица переменных, используемых для описания объектов предметной области, обычно существует в неявном виде как общепризнанный набор характеристик объектов предметной области. Она не соответствует никакому конкретному объекту из предметной области, а только всей предметной области в целом. Для того чтобы получить из нее модель конкретного объекта, эту таблицу надо заполнить конкретными значениями переменных. Таким образом, этап построения семантической модели предметной области можно определить как формализацию зафиксированных в описании логических отношений. Можно представить себе описание предметной области на основе таблицы переменных как n -мерный куб, каждое измерение которого соответствует одной из переменных. Очевидна избыточность данного куба, так как существуют не все объекты, которые могут быть описаны с помощью набора переменных, выбранного для описания объектов предметной области. Таким образом, можно описывать и ограничения, налагаемые на эти объекты. Все это может быть представлено системой уравнений - формализация модели предметной области, где имеет место пересечение множеств семантических свойств, предметных областей объектов, в связи с предполагаемой распределенностью.

Модель предметной области S может быть представлена в виде метабазы, в которой содержится информация по каждому элементу структуры. С другой стороны предметная область W может быть также представлена в виде множества. Введем обозначение: P - множество свойств определяемых связями между элементами указанных выше множеств. Тогда взаимосвязь между ними может быть определена отношением на декартовом произведении $P \times S \times W =$

$$= \{(p_i, r_i, w_i) : p_i \in P, r_i \in S, w_i \in W, i = 1, \dots, n\}$$

Принадлежность элемента , где

$p_i \in P, r_i \in S, w_i \in W, i = 1, \dots, n$ данному отноше-

нию интерпретируется следующим образом: “в объекте модели предметной области r_i содержится информация по свойству p_i объекта предметной области w_j ”.

Поиск информации, соответствующей конкретному элементу модели предметной области r_i в объекте предметной области w_j , сводится к определению отношения $R \subseteq S \times W$. Таким образом, о любой паре

$(r_i, w_j) \in R, w_j \in W, i=1, \dots, n$ можно ска-

зать, что w_j является релевантным и решение задачи определения релевантности элементов множеств S и W , сводится к определению отношения. При этом для любых $r_i \in S, w_j \in W, r_j \in S, w_j \in W, i, j=1, \dots, n$ верно, что если

$r_i \subseteq r_j$, т.е. все элементы объекта

модели предметной области r_i содержатся в объекте модели предметной области r_j и все элементы объекта предметной области w_j со-

держатся в объекте предметной области w_j и $(r_i, w_j) \in R$, то выполняется $(r_j, w_j) \in R$. За исключением крайнего случая, когда отношение R есть само декартово произведение $S \times W$, отношение включает в себя не все возможные кортежи из декартового произведения. Это значит, что для каждого отношения имеется критерий, позволяющий определить, какие кортежи входят в отношение, а какие нет. Таким образом, каждому отношению можно поставить в соответствие некоторое логическое выражение P (предикат отношения R) зависящее от определенного числа параметров (n -местный предикат) и определяющее, будет ли кортеж

принадлежать отношению R (принадлежность кортежа отношению эквивалентна истинности предиката) $(r_j, w_j) \in R \Leftrightarrow \{P\} = \{S, W, R\}$.

В качестве схем информационных объектов рассматриваемой модели могут быть определены независимые бизнес-процессы. Обработка и представление информации, применительно к этапам рассматриваемого

процесса для количественных показателей атрибутов, может производиться применением прямого метода для одного эксперта построения функции принадлежности нечеткого множества, например, предложенным Осгудом методом семантических дифференциалов на основе совокупности оценок по шкалам⁶. Учитывая построение модели в виде объектов, каждый независимый бизнес-процесс можем рассматривать как отдельную (обособленную) часть информационной системы⁷. Учитывая, что система разделена на информационные объекты, объединенные по семантическим правилам взаимодействия, можно декларировать относительную полноту множества учитываемых отношений между элементами системы, которые определяют ее поведение и являются предметом анализа функциональной надежности. Таким образом, отношения между взаимодействующими бизнес-процессами могут быть классифицированы на основе математических правил четкой и нечеткой логик. Концепция построения такой системы отражает фактически современную стратегию так называемых CALS-технологий и может быть рассмотрена как инструмент повышения эффективности и качества, так как полностью соответствует духу и принципам международных стандартов серии ISO-9000.

¹ Бениаминов Е.М. Алгебраические методы в теории баз данных и представлении знаний // Научный мир. М., 2003.

² Хомский Н. Язык и проблема знания // Вестн. МГУ: сб. науч. тр. М., 1996. Вып. 6. С. 157-185.

³ Заболева-Зотова А.В., Камаев В.А. Лингвистическое обеспечение автоматизированных систем. М., 2008.

⁴ Кнут Д. Искусство программирования для ЭВМ. Т. 1. М., 1976.

⁵ Солдатов А.В. Методика создания информационных систем на платформе 1С Предприятие // Научная книга: сб. тр. Воронеж, 2004. Вып. 9. С. 3-132.

⁶ Борисов А.Н., Крумберг О.А., Федоров И.П. Принятие решений на основе нечетких моделей. Примеры использования. Рига, 1990.

⁷ Куликов Г.Г., Антонов В.В. Метод формирования структуры хранилища данных для автоматизированной учетной системы на основе процессного анализа предметной области // Вестн. УГАТУ: сб. науч. тр. Уфа, 2006. Т. 8, № 1 (17). С. 60-67.