

В. В. Антонов, Г. Г. Куликов, Д. В. Антонов

## ПОСТРОЕНИЕ ФОРМАЛЬНОЙ МОДЕЛИ ПРЕДМЕТНОЙ ОБЛАСТИ С ПРИМЕНЕНИЕМ НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ

В статье рассмотрены 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 изменение ее. Доля затрат на получение и переработку информации в экономике и общественных расходах непрерывно возрастает. Однако до настоящего времени 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 описаний модели исследуемого объекта. *Атрибутивная трансляция; предметная область; семантическая модель; методология структурного анализа*

### ВВЕДЕНИЕ

Мы живем на рубеже перехода к информационной эпохе, перехода к внедрению эффективных методов управления сложными и быстротекущими процессами, изменения методов управления не только технологическими процессами, но также методов управления 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 происходят 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, то есть возникают синергетические свойства. При этом важно, что объединение элементов приводит к появлению новых

свойств, которых у них не было, и утрате некоторых свойств элементов, т. е. происходят сложные качественные изменения, которые не всегда могут быть представлены и объяснены.

Но именно благодаря этой особенности рассматриваемая закономерность приводит к возможности отображать системы с неопределенностью в виде формальной четкой модели.

### 1. СОСТОЯНИЕ ВОПРОСА

Перевод условий практической задачи на язык математических моделей всегда был трудным и зачастую приводил к потере трудноформализуемой 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.

Процесс начинается с выделения объектов предметной области и выявления связей между ними. Рассмотрим подробнее процесс выбора объектов. В общем виде функцию выбора можно представить в виде множества альтернативных объектов, выбираемых по некоторому условию, которое в свою очередь может быть представлено в виде совокупности сведений о состоянии объекта и множестве правил выбора.

Пусть  $\sigma$  – совокупность сведений, характеризующих объекты,  $Z$  – множество объектов предметной области,  $z_i \in Z$  – объект из множества объектов. Очевидно, что часть сведений, характеризующих объект, можно представить в виде множества его информационных характеристик  $x_i = \{ \langle A_i, D_i \rangle \mid x_i \subset \sigma \}$ , где  $A_i$  – непустое множество имен свойств (атрибутов)  $i$ -го объекта,  $D_i$  – множество значений соответствующих атрибутов,  $x_i$  – множество информационных характеристик  $i$ -го объекта.

Может быть составлен словарь элементов допустимых значений, подразделенный на классы, что позволяет представить предметную область в виде иерархической структуры. Значения разбиваются на классы объектов, которые взаимодействуют друг с другом на основе правил. Пусть  $\pi$  – множество правил выбора, тогда условия выбора объекта [1] из множества альтернатив могут быть представлены в виде кортежа  $y = \langle \sigma, \pi \rangle$ . На множестве атрибутов могут быть установлены отношения  $G = \{ \bar{G}, \tilde{G} \}$ , кото-

рые делятся на количественные  $\bar{G}$  и качественные  $\tilde{G}$ , и определено множество типов выбора, например  $T = \{ \text{«соответствие»}, \text{«эквивалентность»}, \text{«предпочтение»} \}$ . Тогда любое правило выбора может быть представлено кортежем  $\pi = \langle G, T \rangle$ . Таким образом, совокупность информационных характеристик объекта, установленные отношения и правила установления отношений являются характеристикой объекта  $z_i = \{ x_i, G, \pi \} = \{ \langle A_i, D_i \rangle, \bar{G}, \tilde{G}, T \}$ ,  $i \in N$ . Значения атрибутов могут носить не числовой характер. В частности, в макроэкономических, социологических, маркетинговых, медицинских, правовых хранилищах данных широко используется лингвистическая форма представления данных. В общем случае характеристика каждого объекта  $x_i$  может быть описана соответствующей лингвистической переменной  $\langle A_j, T_j, D_j \rangle$ , где  $T_j = \{ T_1^j, T_2^j, \dots, T_{m_j}^j \}$  – терм-множество лингвистической переменной  $A_j$  (набор лингвистических значений атрибута),  $m_j$  – число значений атрибута;  $D_j$  – (предметная шкала) базовое множество атрибута  $A_j$ . Для описания термов  $T_k^j, k = 1, \dots, m_j$ , соответствующих значениям атрибута  $A_j$ , могут быть использованы нечеткие переменные  $\langle T_k^j, D_j, \tilde{C}_k^j \rangle$ , т. е. значение  $T_k^j$  описывается нечетким множеством

$\tilde{C}_k^j = \left\{ \left\langle \mu_{\tilde{C}_k^j}(d) \mid d \right\rangle \right\}$ , где  $d \in D_j, k = 1, \dots, m_j$ . В результате в качестве нечеткой характеристики объекта  $x_i$  может быть взято нечеткое множество второго уровня

$$\tilde{x}_i = \left\{ \left\langle \mu_{x_i}(a_j) \mid a_j \right\rangle \right\}, \quad (1)$$

где  $\mu_{x_i}(a_j) = \bigcup_{k=1}^{m_j} \left\{ \left\langle \mu_{\mu_{x_i}}(T_k^j) \mid T_k^j \right\rangle \right\}$ ,  $T_k^j \in T_j, a_j \in A_i$ .

Исходя из приведенного, предметную область можно представить в виде многоуровневой среды, состоящей из множества элементов предметной области, множества функций и методов, работающих на этих элементах, множества свойств элементов и отношений между элементами, т. е. в виде онтологии, которая включает в себя описание свойств предметной области и взаимодействия объектов на некотором формальном языке, имеющем логическую семантику. Если система сложная, число факторов велико, то учет всех ее характеристик (компонент) приводит к чрезвычайной сложности. Поэтому в модель приходится вводить лишь ограниченное число, а оставшиеся компоненты учитывать, явно не вводя в модель, но учитывая их влияние как нечеткую реакцию модели на тот или иной выбор альтернативы. Даже для выбранных объектов (свойства которых важны для ИС) размерность остается очень большой. Очевидно, что алгебраическое сравнение компонент невозможно и может быть выполнено с применением методов нечеткой логики.

Таким образом, конечное множество объектов  $Z = \{ z_1, \dots, z_n \}$  предметной области может быть использовано в качестве множества объектов кластеризации. Данное множество описывается конечным множеством атрибутов, каждый из которых количественно представляет некоторое свойство или характеристику элементов рассматриваемой предметной области. Каждый кластер будет иметь свойства OLAP куба. Использование средств нечеткой логики позволяет утверждать, что для каждого объекта в некоторой количественной шкале измерены все значения атрибутов, то есть каждому объекту поставлен в соответствие некоторый вектор, координатами которого являются количественные значения соответствующих атрибутов. Будем характеризовать объекты  $Z = \{ z_1, \dots, z_n \}$ , подлежащие кластеризации векторами параметров (атрибутов)  $y^i = \{ y_1^i, \dots, y_{p_i}^i \}$ , где  $p_i$  – число атрибутов, характеризующих  $i$ -й объект.

Пусть  $F = \{ F_1, \dots, F_k \}$  – множество нечетких кластеров предметной области,  $\mu^F =$

$= \{(\mu_1^1, \dots, \mu_k^1), \dots, (\mu_1^n, \dots, \mu_k^n)\}$  – множество функций принадлежности объектов нечетким кластерам. Тогда сумма степеней принадлежности ко всем кластерам для каждого объекта  $i$  удовлетворяет условию  $\sum_{j=1}^k \mu_j^i = 1$ . Введем дополнительное ограничение, основанное на том, что нечеткие кластеры образуют нечеткое покрытие множества объектов в том и только в том случае, если для каждого кластера  $j$  выполняется  $\sum_{i=1}^n \mu_j^i \geq 1$ .

После задания степеней принадлежности, удовлетворяющих указанным ограничениям, можем вычислить координаты центров кластеров по формуле

$$v^j = \frac{\sum_{i=1}^n (\mu_j^i (y^i)^2 \times \sum_{m=1}^{p_i} y_m^i)}{\sum_{i=1}^n \mu_j^i (y^i)^2}, \quad (2)$$

где  $j = 1, \dots, k$ . Задача сводится к минимизации целевой функции

$$J(Z) = \sum_{m=1}^k \left( \sum_{i=1}^n (\mu_j^i (y^i)^2 \times \sum_{m=1}^{p_i} (y_m^i - v^j)^2) \right). \quad (3)$$

Пусть в процессе кластеризации построено нечеткое покрытие. При рассмотрении очередного объекта на соответствие каждому кластеру  $F = \{F_1, \dots, F_k\}$  и изменении соответствующих функций принадлежности, при возникновении ситуации, когда  $\sum_{j=1}^k \mu_j^{n+1} < 1$ , принимается решение о создании нового нечеткого кластера. При отсутствии информации о значениях каких-либо атрибутов вектора параметров степень нечеткости принадлежности объекта к кластеру определяется максимальным расстоянием от центра кластера.

Понимая под альтернативным решением вариант выбора учитываемых компонент, обозначим через  $X$  множество альтернативных решений. Выбор наиболее предпочтительного решения в каждом конкретном случае может быть осуществлен по совокупности сложных критериев с нормировкой его компонент. На основе данного множества  $X$  сформируем множество упорядоченных пар  $E = X \times X$  альтернативных решений. Обозначим через  $(x, y)$ , где  $x \in X, y \in X$ , пару альтернативных решений, через  $\mu(x, y) \in [0, 1]$  функцию принадлежности нечеткого отношения предпочтения решения  $x$  перед решением  $y$ . Само нечеткое отношение предпочтения можем определить формулой  $P = \{E, \mu(x, y)\}$ . Очевидно, что для каждого  $P$  существует

обратное нечеткое отношение  $P^{-1} = \{E, \mu(y, x)\}$ . Определим степень превосходства решения  $x$  над решением  $y$  в виде  $\Delta(x, y) = \mu(x, y) - \mu(y, x)$ . Очевидно, что  $\Delta(x, y) = -\Delta(y, x)$ . Тогда множество недоминируемых решений может быть определено  $X^{\text{und}} = \{x \mid \Delta(y, x) \leq 0, \forall y \in X\}$ . Учитывая, что нас интересует только множество четко недоминируемых решений, исключим отношение равноценности  $\Delta(x, y) = 0$  и, используя доказательства, приведенные в [5], можем свести решение к выбору альтернатив на основе  $\beta$ -уровневых отношений предпочтения, причем можем ввести одно нечеткое отношение предпочтения, равное  $\sum_{i=1}^m \lambda_i \mu_i(x, y)$ , где  $\lambda_i$  – весовые

коэффициенты важности для нечетких отношений предпочтения. При этом решение сводится к многокритериальной задаче принятия решений, где в качестве критериев эффективности выступают функции принадлежности, и в конечном итоге к вычислению одной функции по формальному алгоритму решения  $\beta$ -уровневых моделей задач принятия решений

$$F(\beta) = \left\{ (x, y) \mid \sum_{i=1}^m \lambda_i \Delta_i(x, y) \geq \beta \right\}. \quad (4)$$

Модель, охватывающая информационную систему, может быть представлена в виде мета-базы, 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 произведении [4]. Рассматривая цели и ограничения в виде симметричных элементов логической схемы, можем достаточно просто сформировать на их основе решение.

## 2. ПРОБЛЕМЫ ПОСТРОЕНИЯ ФОРМАЛЬНОЙ МОДЕЛИ ПРЕДМЕТНОЙ ОБЛАСТИ

В иерархической схеме мы можем выделить горизонтальные и вертикальные бизнес-процессы. Бизнес-процессы, рассматриваемые как функциональная модель реальных процессов, представляют собой структурированное описание заданной последовательности выполняемых бизнес-операций, то есть 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. Вертикальные бизнес-процессы выполняют функции консолидации горизонтальных процессов, учета дополнительно появляющихся атрибутов, описывающих правила взаимодействия информационных характеристик объектов. Они также реализуют учет функциональных правил перевода информации из атрибутов базы данных в хранилища данных. Более того, эти процессы также являются объектами учета.

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. Общеизвестное 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. 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 следующие методологические аспекты:

- при рассмотрении предметной области как части действительности – онтологический;
- 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;
- модель знаний об этой действительности.

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 логических отношений. Теоретической базой для этого служат функциональная, структурная, объектная и концептуальная методологии анализа предметной области, реализуемые соответствующими технологиями программирования. Несмотря на развитость и завершенность перечисленных методологий, не устраненным остается семантический разрыв между содержательными представлениями о предметной области и теми средствами, которые служат для выражения этих представлений в виде формальных спецификаций. Смысл или семантика может определяться разными путями. Отсюда наиболее простым решением представляется организация при помощи машинной компоненты лишь «формальной» обработки данных, оставляя их семантическую интерпретацию за человеком. Однако по мере роста сложности рассматриваемых предметных областей и соответствующих моделей такое решение по многим причинам становится неудовлетворительным. Прежде всего, увеличение объема и сложности моделей предметных областей приводит к тому, что поддержание их удовлетворительного семантического описания становится для человека непосильным. Повышение сложности решаемых задач настоятельно требует учета семантики данных в ходе решения задач. Для сокращения семантического разрыва используют повыше-

ние уровня абстракции модели, что в данном случае связано с проблемой описания семантики. Речь идет об описании свойств отношений между математическими объектами, не зависящими от внутренней структуры объектов – т. е. использование теории категорий. Использование функций в качестве базового примитива позволяет единообразно трактовать конечные и бесконечные области значений. 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. Использование идей метода синтаксически-ориентированной трансляции (основанного на работах Ноэля Хомского) семантический анализ может быть сведен к синтаксическому пошаговому выполнению процессов [3]:

- распознавание структуры;
- построение выходных действий на основе этой структуры.

Д. Кнут формализовал подобные идеи, введя понятие «атрибутивной трансляции» [2]. Символам грамматики ставится в соответствие конечное число семантически определенных атрибутов, значения которых могут быть числовыми, символьными, матрицами или какими-либо признаками. Правила грамматики определяют связи этих атрибутов, а вычисление значений определяется деревом вывода. 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. Получается, что ранее выделенные объекты оказываются в более привилегированном положении по отношению к последующим объектам предметной области, поскольку модели этих объектов конструируются при помощи изменения модели уже известных объектов. При этом последовательность выбора объектов есть воля случая. При попытке выбора объекта, который наилучшим образом подходит в качестве первого, выясняется, что наиболее удобно использовать в его качестве образ некоторого идеализированного, средне-статистического объекта этой предметной области, модель которого заменена на набор переменных, описывающих объекты предметной области. Выбор набора переменных для описания объектов предметной области и выбор допустимых значений для этих переменных в значительной степени произволен. Однако именно этот выбор в дальнейшем будет определять границы применимости ее модели.

Пусть  $W$  – исследуемая предметная область объектов  $W = \{w_1, \dots, w_n\}$ . Часто в моделях рассуждений присутствуют нечеткие понятия, однако любая информация, записанная в каком-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 функцией в обычном понимании, а определяет лишь возможное состояние предметной области одного объекта на основе отличия от другого. Т. е. 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  $r_{i+1} = F_{i+1}^M(w_{i+1})$ , получаем, что  $r_{i+1} = F_{i+1}^M(F_i^W(w_i))$ . 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 предметной 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. 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.

Модель предметной области  $S$  может быть представлена в виде метабазы, в которой содержится информация по каждому элементу структуры. С другой стороны, предметная об-

ласть  $W$  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

$$P \times S \times W = \left\{ (p_i, r_i, w_i) : p_i \in P, \right. \\ \left. r_i \in S, w_i \in W, i = 1, \dots, n \right\}. \quad (5)$$

Принадлежность элемента  $z_i = (p_i, r_i, w_i)$ , где  $p_i \in P$ ,  $r_i \in S$ ,  $w_i \in W$ ,  $i = 1, \dots, n$ , данному отношению интерпретируется следующим образом: «в объекте модели предметной области  $r_i$  содержится информация по свойству  $r_i$  объекта предметной области  $w_i$ ».

Поиск информации, соответствующей конкретному элементу модели предметной области  $r_i$  в объекте предметной области  $w_i$ , сводится к определению отношения  $R \subseteq S \times W$ .

Таким образом, о любой паре  $(r_i, w_i) \in R$ :  $r_i \in S$ ,  $w_i \in W$ ,  $i = 1, \dots, n$  можно сказать, что  $w_i$  является релевантным  $r_i$  и решение задачи определения релевантности элементов множеств  $S$  и  $W$  сводится к определению отношения  $R \subseteq S \times W$ . При этом для любых  $r_i \in S$ ,  $w_i \in W$ ,  $r_j \in S$ ,  $w_j \in W$ ,  $i = 1, \dots, n$  верно, что если  $w_i \subseteq w_j$  и  $r_i \subseteq r_j$ , то есть все элементы объекта модели предметной области  $r_i$  содержатся в объекте модели предметной области  $r_j$  и все элементы объекта предметной области  $w_i$  содержатся в объекте предметной области  $w_j$  и  $(r_i, w_i) \in R$ , 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  $P$ , которое является предикатом отношения  $R$ , зависящее от определенного числа параметров ( $n$ -местный предикат) и определяющее, будет ли кортеж  $(r_j, w_j)$  принадлежать отношению  $R$ , что эквивалентно истинности предиката  $\{P\} = \{S, W, R\}$ .

### 3. ПОСТРОЕНИЕ ФОРМАЛЬНОЙ МОДЕЛИ АБСТРАКТНОГО БИЗНЕС-ПРОЦЕССА

Наиболее наглядным следует считать описание формального языка в виде синтаксических диаграмм, представляющих собой графическое представление форм языка. Выбор мето-

дологии структурного анализа напрямую зависит от специфики предметной области, для которой создается модель. Для построения модели системы на основе ее описания с использованием методологии IDEF0 требуется адекватная целям моделирования интерпретация стандарта. Так, возможно представление системы в виде множества взаимодействующих взаимосвязанных функций, что позволяет рассматривать функции независимо от объектов, которые их выполняют. Все это позволяет отделить проблемы анализа и проектирования от проблем реализации. При этом модель представляет собой набор иерархически упорядоченных диаграмм, каждая из которых описывает определенную функцию и состоит из нескольких взаимодействующих взаимосвязанных подфункций, связанных между собой как горизонтальными, так и вертикальными связями. Это позволяет не просто описать структуру функций, но и их взаимодействие, придающее совокупности функции системные свойства. Система может быть представлена в виде множества процессов  $E$ , осуществляющих преобразования элементов системы. Рассмотрим модель процесса  $E_i$  с  $N$  входами,  $K$  выходами,  $L$  управлениями и  $J$  механизмами абстрактного бизнес-процесса.

Пусть  $I_i$  – множество  $\{i_1, \dots, i_N\}$  входов процесса  $E_i$ ,  $O_i$  – множество  $\{o_1, \dots, o_K\}$  выходов процесса  $E_i$ ,  $C_i$  – множество  $\{c_1, \dots, c_L\}$  управлений процесса  $E_i$ ,  $M_i$  – множество  $\{m_1, \dots, m_J\}$  механизмов процесса  $E_i$ ,  $F_i$  – множество  $\{f_1, \dots, f_P\}$  взаимосвязанных функций преобразующих входы в выходы процесса  $E_i$ . Тогда процесс может быть представлен в следующем формальном виде  $E_i = \langle I_i, C_i, O_i, M_i, F_i \rangle$ . При этом, взаимодействия процесса  $E_i$  с процессом  $E_j$  исчерпываются следующим множеством отношений:

Выход-вход;  $G_{OI}$  – множество отображений выходов  $O$  на входы  $I$ ;

Выход-управление;  $G_{OC}$  – множество отображений выходов  $O$  на управления  $C$ ;

Выход-механизм;  $G_{OM}$  – множество отображений выходов  $O$  на механизмы  $M$ .

Пусть  $E$  – множество  $\{E_1, \dots, E_z\}$  процессов системы,  $G$  – множество отношений процессов  $\{E_1, \dots, E_z\}$ , определяющих постоянные взаимосвязи и динамические взаимодействия компонентов системы, определяемое на следующем пространстве значений  $G = \langle G_{OI}, G_{OC}, G_{OM} \rangle$ . IDEF предусматривает дополнительное описание полной иерархии объектов системы посредством формирования глоссария для каждой диаграммы модели и объединения этих глоссариев в словарь стрелок –  $S$ , который является основ-

ным хранилищем полной иерархии объектов системы.

Учитывая приведенное выше, модель системы может быть представлена в следующем формальном виде:  $C = \langle E, G_{OI}, G_{OC}, G_{OM}, S \rangle$ . Например, выход  $k_1$  процесса  $i$  является входом  $k_2$  для  $j$ -го процесса:  $v_i^{k_1} = G_{OI}(e_i)$ ,  $i_j^{k_2} = v_i^{k_1}$ . Отображение может иметь нечеткий характер. В этом случае возможно получение множества допустимых вариантов отображений.

Весь процесс построения информационной системы будет характеризоваться конечным числом взаимодействий бизнес-процессов, что позволяет представить информационную систему в виде управляемой инвариантной по времени детерминированной системы с конечным числом состояний  $q_i$ , принадлежащих заданному конечному множеству возможных состояний  $q_i \in Q$ ,  $i = 1, \dots, n$ . Сам процесс построения  $w_i$ , в каждый детерминированный момент времени, будет являться элементом множества процессов построения  $W$ . Тогда динамика информационной системы при переходе из одного состояния в другое будет описываться функцией отображения  $F: Q \times W \rightarrow Q$ , т. е.  $f(q_i, w_i)$  есть следующее состояние информационной системы после выполнения процесса построения ( $w_i$ ), т. е.  $q_{i+1} = f(q_i, w_i)$ , где  $i = 1, \dots, n$ .

При этом присутствуют правила ограничения  $C_i = \{c_{ij}: j \in I\}$ , которые являются нечетким множеством в  $W$  с функцией принадлежности  $\mu_{C_i}(w_i)$ . Т. е. цель и правила взаимодействия рассматриваются как нечеткие множества в одном и том же пространстве. Эта симметрия позволяет не делать между ними различия при формировании решения, нечеткие понятия формализуются в виде нечетких и лингвистических переменных, а нечеткость действий в виде нечетких алгоритмов. Таким образом, неопределенность знаний на каждом этапе построения информационной системы складывается из двух компонентов: первый определяется видом построенной части формальной модели и второй – это оставшаяся неопределенность тех подсистем исходной физической системы, которые остаются не исследованными. В итоге задача сводится к количественной оценке этих величин и демонстрационному их связи с исходной неопределенностью физической системы.

В качестве схем информационных объектов рассматриваемой модели могут быть определены независимые бизнес-процессы. Учитывая построение модели в виде объектов, каждый

независимый бизнес-процесс можем рассматривать как отдельную (обособленную) часть информационной системы. Любое информационное взаимодействие между сложными системами реализуется последовательно на физическом, синтаксическом и семантическом уровнях взаимодействия. Учитывая, что наша 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. Учитывая выше сказанное, путем реализации процессного мониторинга на основе формализации сложившихся технологий, представляется возможным представление бизнес-процессов в виде совокупности взаимодействующих вертикальных и горизонтальных семантически определенных и формализованных объектов:

- по горизонтали – автономные бизнес-процессы, согласно организационным звеньям подразделений;

- по вертикали – бизнес-процессы, которые взаимодействуют по семантическим правилам согласно функциональному назначению.

В результате получим иерархическую структуру с условным разделением на «объекты» – для каждого из которых возможен дифференцированный подход:

- зафиксирован перечень организационных звеньев и их иерархия;

- определен перечень функций, которые выполняются в организации, и их взаимосвязи по иерархии – функциональные модели;

- определено, каким образом функции закрепляются за организационными звеньями или «кто за что» отвечает.

Перечень всех функций дает положение об этом организационном звене, которое традиционно называют положением о подразделении.

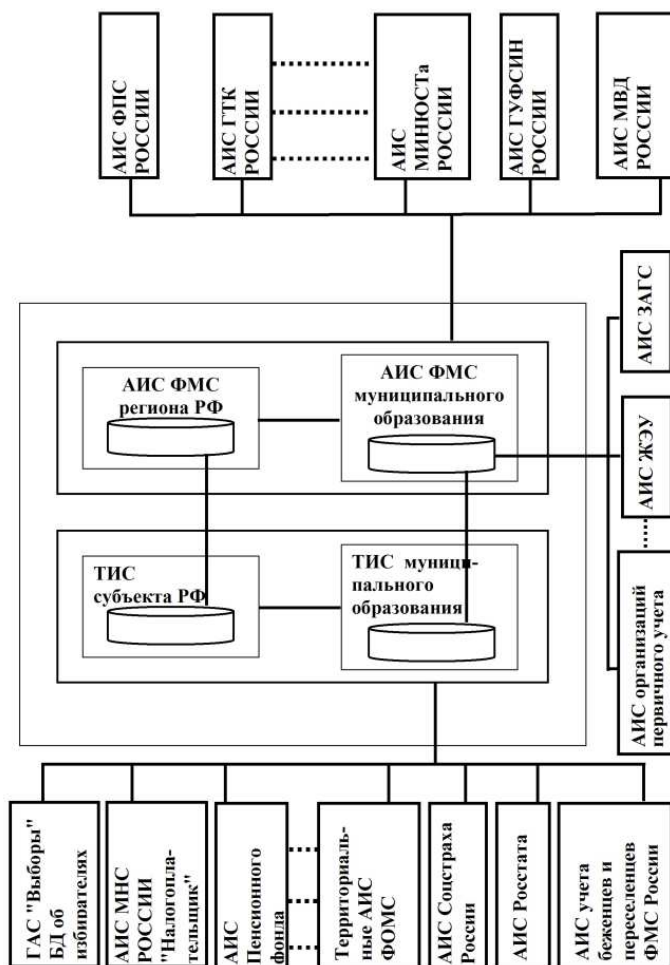
Внутри каждого «объекта», при введении автоматизации, бизнес-процессы легко преобразуются в сквозные, а все уровни соединяются

между собой в единую функциональную систему. При этом в основе данной системы лежит набор интегрированных информационных моделей, состоящий из самого жизненного цикла системы и выполняемых в его ходе бизнес-процессов. Единая интегрированная модель описывает объект достаточно полно и может быть использована в роли единого источника информации для любых выполняемых в ходе жизненного цикла процессов. Системная информационная поддержка осуществляется в интегрированной информационной среде, определяемой как совокупность распределенных баз данных.

Примером учета может служить регистрационный учет населения. С точки зрения развития экономики страны человеческие ресурсы играют двуединую роль. С одной стороны – это активная часть производительных сил, прямой и косвенный источник формирования доходной части бюджета государства. С другой стороны – это объект прямых и косвенных затрат бюджета. В связи с этим для формирования обоснованных планов социально-экономического развития страны необходимо иметь достоверную информацию о количественном и качественном составе населения территориальных образований, входящих в состав России, а также о факторах, влияющих на демографические процессы.

Учетные процедуры проводятся одновременно различными ведомствами в соответствии с решаемыми ими задачами, в связи с этим охватывают различные сегменты населения (рисунок). Всего в Российской Федерации в настоящее время существует более 18 ведомственных баз данных, содержащих различное количество записей о физических лицах. Перечень характеристик лица в системах персонального учета зависит от возможности получать точную и своевременную информацию об изменениях включенных в реестр признаков. В связи с этим многие системы персонального учета содержат разумный минимум необходимых и своевременно актуализируемых показателей, а возможности по сбору дополнительных сведений расширяются постепенно, по мере создания отраслевых подсистем, содержащих другие аспекты персональной информации, а также при внедрении процедур межведомственного информационного обмена. Так, при многократном прохождении информации из разных источников автоматически могут быть сформированы показатели качества (подтверждения тех или иных параметров).





Анализ состояния предметной области учета персональных данных категорий населения

### ЗАКЛЮЧЕНИЕ

Предложенный метод позволяет, поэтапно выделяя объекты предметной области, устанавливая связи на основе семантически определенных атрибутов, строить модель предметной области, снижая неопределенность знаний на каждом этапе. Весь анализ произведен на верхнем уровне абстракции, где в качестве информационных объектов рассматриваемой модели определены независимые бизнес-процессы семантических блоков, что определяет возможность использования отдельных методов и алгоритмов при разработке и исследовании широкого круга информационных задач.

### СПИСОК ЛИТЕРАТУРЫ

1. **Бениаминов Е. М.** Алгебраические методы в теории баз данных и представлении знаний. М.: Научный мир, 2003. 184 с.
2. **Кнут Д.** Искусство программирования для ЭВМ. Т. 1. М.: Мир, 1976. 736 с.
3. **Хомский Н.** Язык и проблема знания // Вестник МГУ. 1996. Вып. 6. С. 157–185.

4. **Айзерман М. А., Алесекров Ф. Т.** Выбор вариантов. Основы теории. М.: Наука, 1990. 240 с.

5. **Погонин В. А.** Модели диспетчерского управления роботами // Информационные процессы и управление. 2006. № 1. С. 45–57.

### ОБ АВТОРАХ

**Антонов Вячеслав Викторович**, доц. каф. АСУ. Дипл. математик (БГУ, 1979). Канд. техн. наук по управл. в соц. и экон. системах (УГАТУ, 2007). Иссл. в обл. автоматиз. информ. систем.

**Куликов Геннадий Григорьевич**, проф., зав. той же каф. Дипл. инж. по автоматиз. машиностроения (УАИ, 1971). Д-р техн. наук по системн. анализу, автоматическ. упр. и тепловым двигателям (УАИ, 1989). Иссл. в обл. АСУ и упр. силовыми установками ЛА.

**Антонов Дмитрий Вячеславович**, асп. той же каф. Дипл. инженер по автоматизир. системам обработки информации и управления (УГАТУ, 2010). Готовит дис. в обл. построения информационных систем.