

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

ФОРМАЛИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ С ПРИМЕНЕНИЕМ ИНСТРУМЕНТОВ, ПОДДЕРЖИВАЮЩИХ СТАНДАРТЫ

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. На основе модели абстрактного бизнес-процесса рассмотрим возможность поэтапной автоматизации и построения информационных систем. *Нечеткая логика; информационные системы; предметная область; BPMN*

Мир, в котором мы живем, является миром со слабыми причинно-следственными связями. Именно благодаря слабой связанности мира, мы можем выделить в нем отдельные системы, а в них подсистемы. В противном случае весь мир представлял бы собой одну чрезвычайно сложную систему. Построение информационной системы, переход из одного состояния в другое происходит по определенным закономерностям. Причем закономерности переходов не всегда могут задаваться в четком виде, т. е. система обладает поведением, определяемым закономерностями [6]. Закономерность целостности проявляется при каждом состоянии системы. Благодаря этому качеству 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. То есть происходят сложные качественные изменения, которые не всегда могут быть представлены и объяснены. И благодаря именно этой особенности мы можем отображать системы с неопределенностью в виде формальной четкой модели. При этом ставится следующая задача:

- переход от управления традиционными базами данных к управлению обширными коллекциями структурированных, полуструктурированных и неструктурированных данных;
- переход к управлению данными, которые распределены по многим репозиториям (то

есть практически к управлению пространствами данных);

- осуществление интеграции данных в условиях их семантической неоднородности, разнотемповости, неполноты и неточности.

В некоторых случаях границы между пространствами данных могут быть плавающими, поэтому все виды связей между участниками должны быть формализованы, а семантическая интеграция развиваться во времени только там, где требуется.

ПРОБЛЕМЫ ПОСТРОЕНИЯ ФОРМАЛЬНОЙ МОДЕЛИ ПРЕДМЕТНОЙ ОБЛАСТИ

Вопрос соотношения реального объекта и его модели адекватен вопросу соотношения некоторого объекта и нашего знания о нем, то есть в модели присутствует неопределенность, которую необходимо учитывать при переносе выводов на реальный объект. Упомянутая «неопределенность» обладает сложной структурой понятия [4, 7]. Причем разные типы неопределенности обрабатываются разными средствами поддержки обработки информации (рис. 1).

Типичный подход к интеграции информации, основанный на построении хранилищ и витрин данных на основе извлечения операционных данных, их трансформации к единой схеме, в данном случае не всегда приемлем.

Теоретической базой для этого служат функциональная, структурная, объектная и концептуальная методологии анализа предметной области, реализуемые соответствующими технологиями программирования.

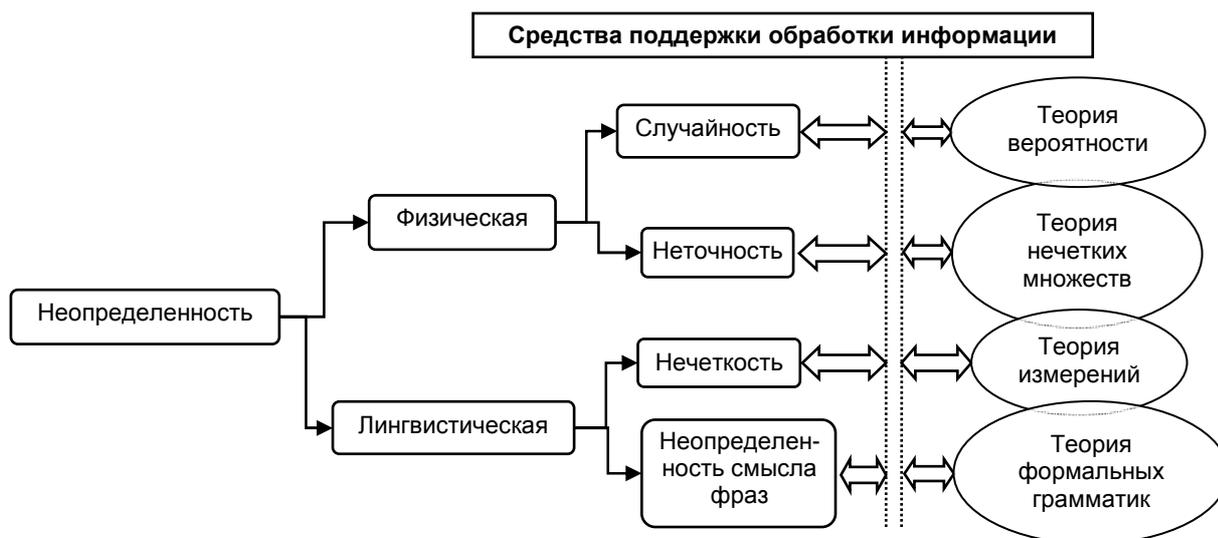


Рис. 1. Средства поддержки обработки информации

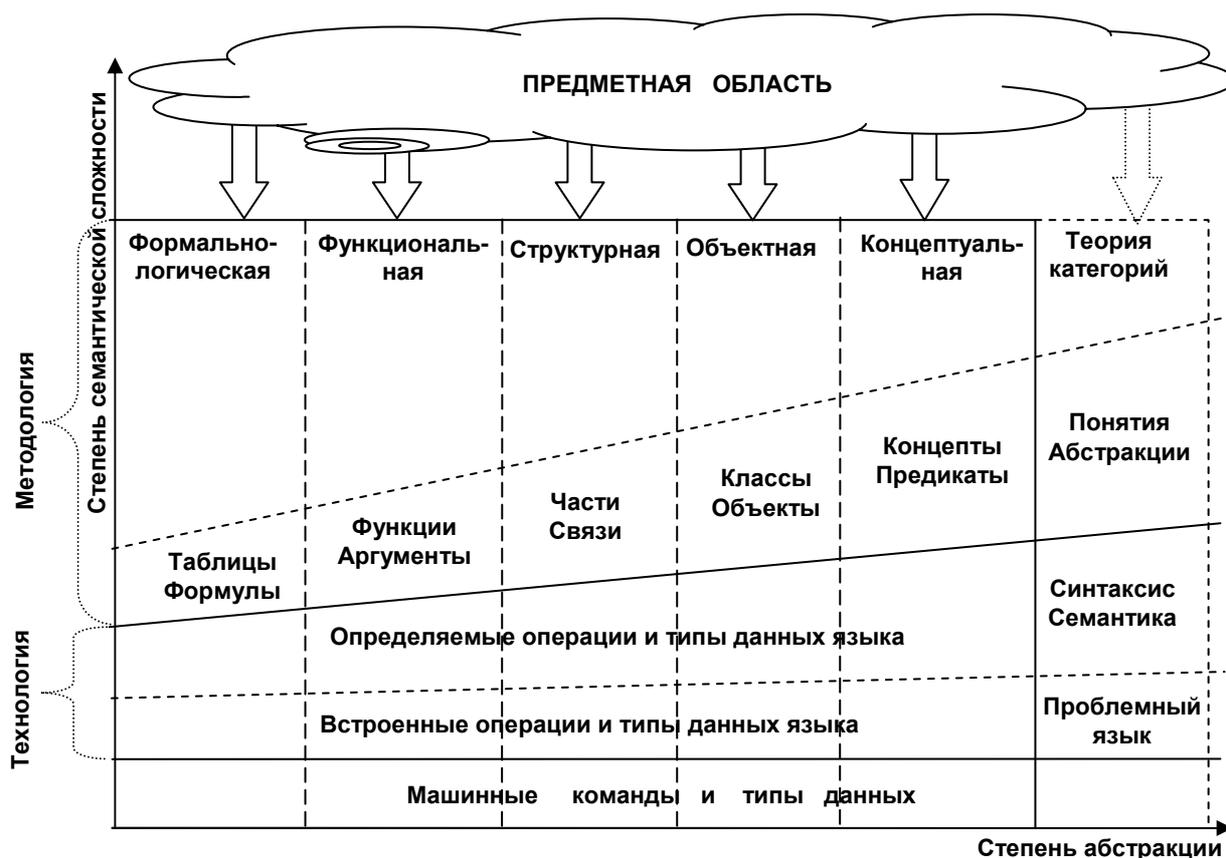


Рис. 2. Теоретическая база методологий анализа предметной области

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 (рис. 2).

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 [2, 8]. 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 [1]. Выявление места несоответствия модели и реальности возможно только в случае полной «семантической управляемости» модели, т. е. осознания разработчиком способов ее применения и характеров получаемых результатов. В ряде ПО постановка задач носит так называемый некорректный характер. При решении таких задач должны учитываться многозначность, неполнота, нечеткость и другие подобные особенности знаний о ПО, не говоря уже о динамической составляющей в связи со всем перечисленным. Однако даже в этих случаях все перечисленные характеристики относятся к характеру связи модели ПО с отображаемой ею реальностью. Сама же модель и тем более способ ее реализации не могут носить нечеткого и т. п. характера. В противном случае становится невозможной успешная реализация системы.

Для сокращения семантического разрыва необходимо повышение уровня абстракции модели, что напрямую связано с проблемой описания семантики [3, 4]. Речь идет об описании свойств отношений между математическими объектами, не зависящими от внутренней структуры объектов – т. е. использование теории категорий. В частности, теория категорий неявно предполагает многоуровневое рассмотрение предметной области. Кроме того, исполь-

зование функций в качестве базового примитива позволяет единообразно трактовать конечные и бесконечные области значений.

Предметную область любой физической системы можно рассматривать как некоторую пространственно-временную область с взаимодействующими объектами-процессами. При попадании нескольких объектов в область взаимодействия, происходит процесс их взаимодействия [4]. 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. Эти условия могут быть представлены в виде совокупности сведений о состоянии объекта и множестве правил выбора (рис. 3).

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 [7].

Имеем два различных класса моделей: модель действительности; модель знаний про эту действительность.

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.

При использовании идей метода синтаксически-ориентированной трансляции (основанного на работах Ноэля Хомского) [8] семантический анализ может быть сведен к синтаксическому пошаговому выполнению процессов: распознавание структуры; построение выходных действий на основе этой структуры.

Таким образом, предлагаемый метод позволяет ограничиться множеством цепочек, которые можно определить некоторым точным образом. То есть можем говорить о некотором формальном языке, заданном в виде множества.

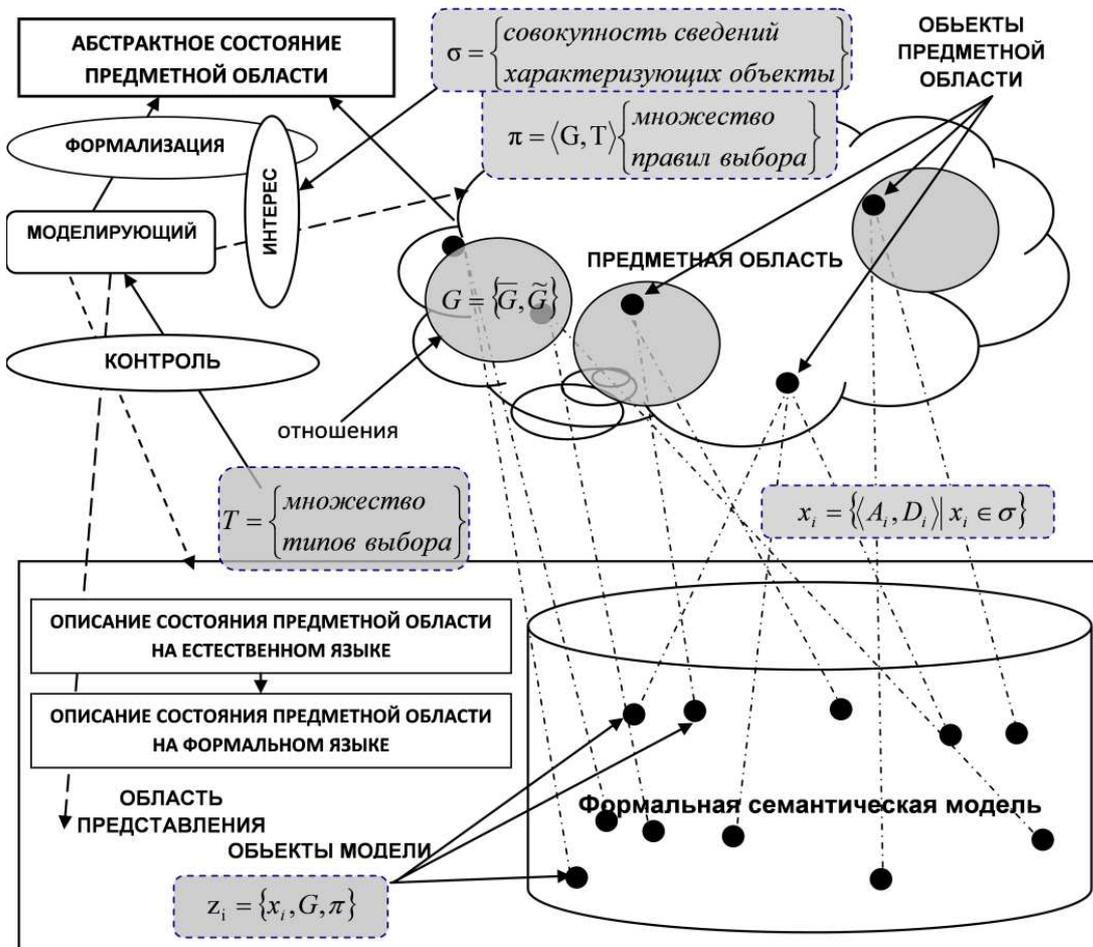


Рис. 3. Моделирование предметной области

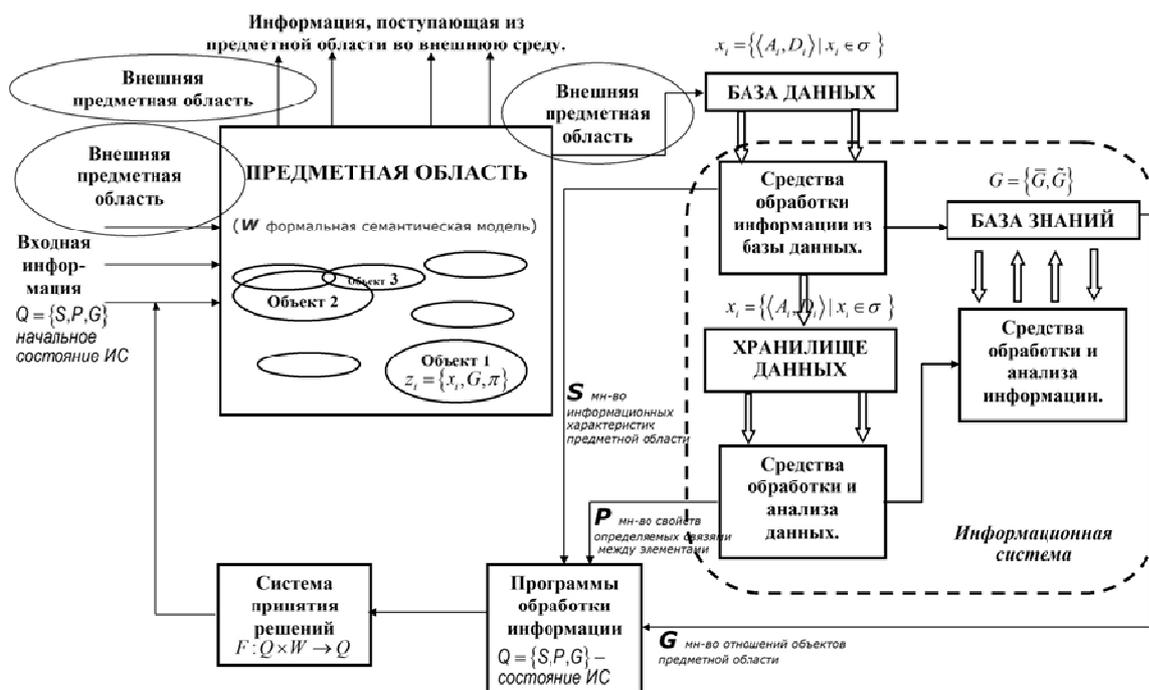


Рис. 4. Отображение предметной области в информационной системе

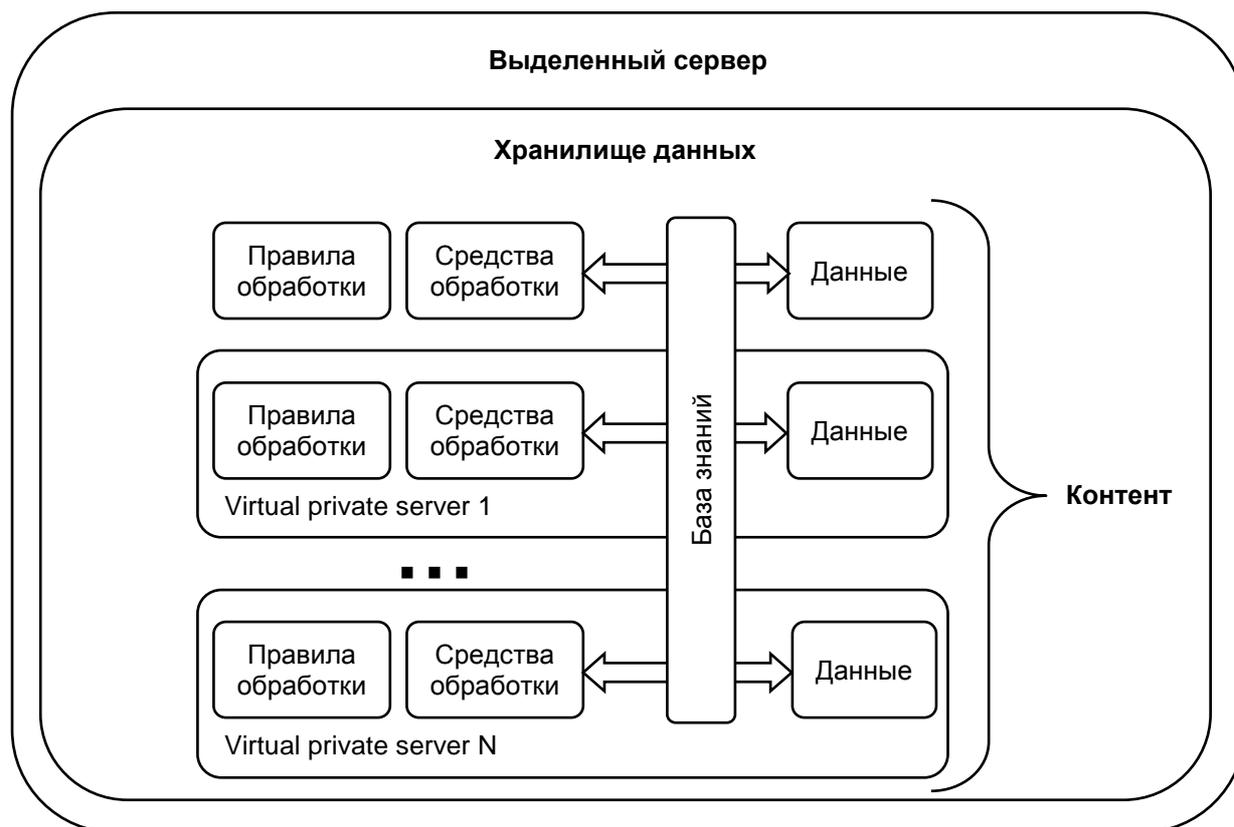


Рис. 6. Понимание контента в рамках виртуализации хранилищ данных

Таким образом, конечное множество объектов 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. Эта проблема известна в математике как проблема покрытия информационного пространства с информационными объектами.

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 [4, 7].

Введя функцию принадлежности каждого объекта к нечетким кластерам, учитывая, что нечеткие кластеры образуют нечеткое покрытие множества объектов (по построению) задачу можно свести к минимизации целевой функции

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

где $Z = \{z_1, \dots, z_n\}$ – объекты кластеризации, $y^i = \{y_1^i, \dots, y_{p_i}^i\}$ – вектор параметров, $\mu^F = \{(\mu_1^1, \dots, \mu_k^1), \dots, (\mu_1^n, \dots, \mu_k^n)\}$ – множество функций принадлежности объектов нечетким кластерам, v^j – координаты центров кластеров.

При выполнении нечетких выводов используются нечеткие соответствия, заданные между областью определения и областью значения в виде прямого произведения R . При этом можем утверждать, что система отображает семейства нечетких множеств друг на друга. Используя основное правило вывода традиционной логики *modus ponens*, согласно которому мы судим об истинности высказывания B по истинности высказывания A , можем применить один из способов формализации соответствия, и применяя разные формализации нечеткого условного высказывания «Если x есть A , то y есть B » можем получить из одной посылки разные выводы, имея возможность выбора конкретного обобщения для решения в зависимости от свойств предметной области [7]. Используемая

математическая модель корреспондируется с положениями теории категорий, которая является основой для теории множеств.

Учитывая, что нас интересуют только множество четко недоминируемых решений, исключив отношение равноценности $\Delta(x, y) = 0$, и используя доказательства, приведенные в [5], решение сводится к выбору альтернатив на основе β -уровневых отношений предпочтения, причем можно ввести одно нечеткое отношение предпочтения равное $\sum_{i=1}^m \lambda_i \mu_i(x, y)$, где λ_i – весо-

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

$$F(\beta) = \left\{ (x, y) \mid \sum_{i=1}^m \lambda_i \mu_i(x, y) \geq \beta \right\}. \quad (2)$$

В иерархической схеме мы можем выделить горизонтальные и вертикальные бизнес-процессы. Горизонтальные бизнес-процессы выполняют разделение реальных объектов внутри класса и для целей учета атрибутов объектов. 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.

Чтобы создать информационную систему, требуется иерархическая организационная структура (рис. 7).

Стрелки Обратной связи показывают, как система разработана, функционирует и корректируется. Бизнес-процессы выделены в определенные классы в соответствии с территориальной принадлежностью (TD). При этом есть взаимосвязь между классами (некоторые объекты могут принадлежать нескольким классам). Это нечеткое свойство определено на функциональном уровне (FD). Используя приведенные нечеткие правила, можем формировать новые кластеры с этими нечеткими объектами. Размер нового создаваемого кластера будет зависеть от величины нечеткого значения β .

База знаний содержит полностью определенные и нечеткие правила, база данных сохраняет только значения атрибутов объектов. В процессе учета и создания информационных объектов мы увеличиваем количество информации. В результате кластеризации можно уменьшить ненужную избыточную информацию и количество информационных объектов. Аналитическая обработка хранилища выполняется в аналитическом центре с обратной связью к более низким уровням иерархии.

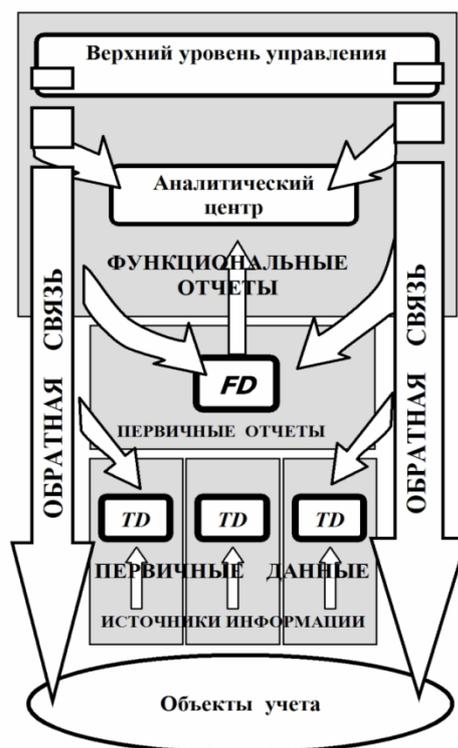


Рис. 7. Статистический / аналитический учет

Для информационной системы можно применить процессный подход, основанный на формальных моделях жизненного цикла (ЖЦ) систем. При этом следует обратить внимание на стандарт ISO/IEC 15288, который является рамочным (Framework) в том плане, что не задает общих требований к реализации процессов, связанных с разработкой и поддержкой жизненного цикла систем и используется в качестве методологической основы для организации этих процессов с необходимой конкретизацией для конкретной области деятельности, при этом жизненный цикл систем представляется в виде «дерева» процессов. С точки зрения процессного подхода, стандарт ISO/IEC 15288 представляет собой обобщенную систему мышления и существенный запас знаний о продуктах и услугах,

явное введение в организационную практику современных парадигм создания продуктов и сервисов через предписание следовать международным стандартам. Стандарт обеспечивает техническую основу для увязывания потребностей рынка с созданием продукции и предоставлением услуг. При этом модель жизненного цикла интерпретируется как концептуальное сегментирование определения нужды в системе, ее создания как продукта или сервиса, ее использования и эволюции. В целях организации планирования, обеспечения, функционирования и поддержки интересующей системы модель жизненного цикла системы сегментирована на стадии, что обеспечивает упорядоченное продвижение системы через установленные барьеры принятия решений с целью уменьшить риск и убедиться в достаточности прогресса.

При этом обрабатываемые данные могут быть представлены в виде конечных последовательностей знаков и цифр, взятых из некоторого конечного алфавита, а сам процесс обработки в виде осуществления преобразования входных данных в выходные посредством разделения на части, выполнения над которыми некоторой последовательности операций (функций или процессов) для получения результатов и формирования на их основе выходных данных. Могут быть использованы различные подходы, структурирующие предметную область и моделирующие ее, например, с помощью описания функциональной структуры. Однако в любом случае, при информационном подходе для формализации предметной области, первичными являются категории объектов и отношения между ними, т.е. формально система отношений может быть представлена множеством объектов предметной области и множеством отношений между ними, а совокупность информационных характеристик процесса $\{ \langle A_i, D_i \rangle, i = 1, \dots, n \}$, установленных отношений $G = \{ \bar{G}, \tilde{G} \}$ и правил установления отношений $\pi = \langle G, T \rangle$ может быть использована для формального определения процесса в виде следующего кортежа компонентов:

$$Z = \{ \langle A_i, D_i \rangle, \{ \bar{G}, \tilde{G} \}, \langle \{ \bar{G}, \tilde{G} \}, T \rangle \} \quad i \in N. \quad (3)$$

Анализ результатов, полученных в [1], позволяет сделать заключение, что предлагаемые методы проектирования и исполнения бизнес-процессов могут быть применимы для разработки всей необходимой для управленческого, проектного и производственного менеджмента номенклатуры бизнес-процессов. Для практиче-

ских нужд моделирования бизнес-процессов, связи между элементами системы могут иметь различный характер, свойства и атрибуты, то есть тоже могут рассматриваться как объекты системы. Например, процессы соглашения PS состоят из процесса приобретения PS_{pr} и процесса поставки PS_{po} . При этом процесс приобретения состоит из цели процесса приобретения $PS_{pr}^1 = \{ pr_1^1 \}$, результата процесса приобретения $PS_{pr}^2 = \{ pr_1^2, \dots, pr_7^2 \}$, деятельности в процессе приобретения $PS_{pr}^3 = \{ pr_1^3, \dots, pr_8^3 \}$. Процесс поставки состоит из цели процесса поставки $PS_{po}^1 = \{ po_1^1 \}$, результата процесса поставки $PS_{po}^2 = \{ po_1^2, \dots, po_7^2 \}$, деятельности в процессе поставки $PS_{po}^3 = \{ po_1^3, \dots, po_8^3 \}$. Имеет место формула, которая определяет отношение между организациями в виде упорядоченного множества:

$$PS = \langle PS_{pr}, PS_{po} \rangle = \langle \{ PS_{pr}^1, PS_{pr}^2, PS_{pr}^3 \}, \{ PS_{po}^1, PS_{po}^2, PS_{po}^3 \} \rangle. \quad (2)$$

Необходимость располагать адаптивными бизнес-процессами, которые могут быть тонко настроены и оптимизированы на соответствие, с одной стороны, нормативной базе, а с другой – изменяющимся условиям бизнеса, привело к системам управления бизнес-процессами. Конструктивной формой описания процессов следует считать также описание в виде синтаксических диаграмм, представляющих собой графическое представление. Из наиболее известных и широко применяемых можно назвать методику структурного анализа и проектирования SADT и основанную на ней IDEF0, а также относительно новую методологию описания бизнес-процессов Business Process Modeling Notation (BPMN).

В качестве примера рассмотрим стадию жизненного цикла – разработки. Стадия разработки Z_2 начинается с детального технического уточнения системных требований и проектных решений, их преобразования в один или несколько реализуемых продуктов. Целью стадии разработки является создание системы, которая удовлетворяет требованиям приобретающей стороны и может быть создана, испытана, оценена, применена по назначению, поддержана при применении и списана. Результаты стадии посредством обратной связи могут быть использованы также для рекурсивного использования внутри самой стадии. Данная стадия может пересекаться со стадиями производства.

Исследуемые параметры:

1. Результат стадии замысла z_1 .
2. Информация о технических и программных средствах и интерфейсах операторов $ZJ_2 = \{zj_2^1, \dots, zj_2^{m_2}\}$.
3. Информация о требованиях к средствам производства, обучения и поддержки $ZJ_3 = \{zj_3^1, \dots, zj_3^{m_3}\}$.
4. Информация из результатов стадии производства ZK_4 .
5. Результаты стадии, используемые для рекурсивного анализа $ZJ_4 = \{zj_4^1, \dots, zj_4^3\}$:
 - соответствие критериям перехода на следующую стадию;
 - санкционирование перехода на стадию производства;
 - подтверждение того, что рассматриваемая система соответствует всем требованиям правообладателей и системным требованиям, может быть запущена в производство, применяться по назначению и поддерживаться в процессе применения, а также может переводиться в категорию непригодных к применению (списываться) и является эффективной по стоимости для правообладателей.
6. Результаты стадии, используемые для последующих стадий – рассматриваемая система или прототип рассматриваемой системы, усовершенствованные обеспечивающие системы или имеющиеся обеспечивающие системы, документация и оценки стоимости последующих стадий и т. д. $ZW_2 = \{zw_2^1, \dots, zw_2^{10}\}$.

Имеет место формула:

$$Z_2 = \langle ZW_1, ZJ_2, ZJ_3, PP, PPR, TP, ZJ_4, ZK_4, ZW_2, \{\bar{G}_2, \tilde{G}_2\}, \langle \{\bar{G}_2, \tilde{G}_2\}, T_2 \rangle \rangle, \quad (3)$$

где $\bar{G}_2, \tilde{G}_2, T_2$ определены в соответствии с формулой (1) для стадии Z_2 . Типовое описание процессов стадии разработки может быть представлено в виде диаграммы, в форме близкой к BPMN, приведенной на рис. 8.

Используя аналогичные описания для других стадий жизненного цикла системы при проектировании и автоматизации процессов можно ясно и лаконично описать:

- логику обработки событий и исключений;
- появление параллельных потоков работ и их синхронизацию;
- взаимодействие процессов и исполнителей.

Все это позволяет определить логические правила для формирования функционального взаимодействия бизнес-процессов как на стадиях ЖЦ проекта, так и в пределах отдельных стадий.

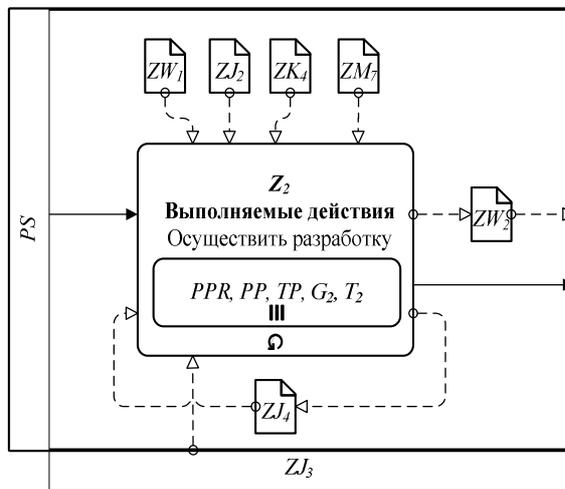


Рис. 8. Диаграмма типового описания процессов стадии разработки в нотации BPMN

В качестве примера сформируем структуры формальной семантической модели жизненного цикла проекта в общем виде в аналитической форме (4) и интерпретируем ее в форме функциональной модели процесса формирования ЖЦ проекта верхнего уровня в нотации BPMN (рис. 9).

$$EC = \left\langle \left\{ Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, PP, PPR, TP, \{\bar{G}, \tilde{G}\}, \langle \{\bar{G}, \tilde{G}\}, T \rangle \right\}, PS \right\rangle. \quad (4)$$

Применение предложенной методики описания БП в условиях нечеткой исходной информации может быть эффективным в случае большого количества независимых входных и выходных параметров, характеризующих состояние системы.

При таком подходе описание бизнес-процессов логично вписывается в контекст описания бизнеса организации в целом и, что самое важное, оно может быть создано исходя из целей организации с последовательной детализацией, уточнением основных процессов и согласовано с описаниями других аспектов бизнеса, например, организационной структурой.

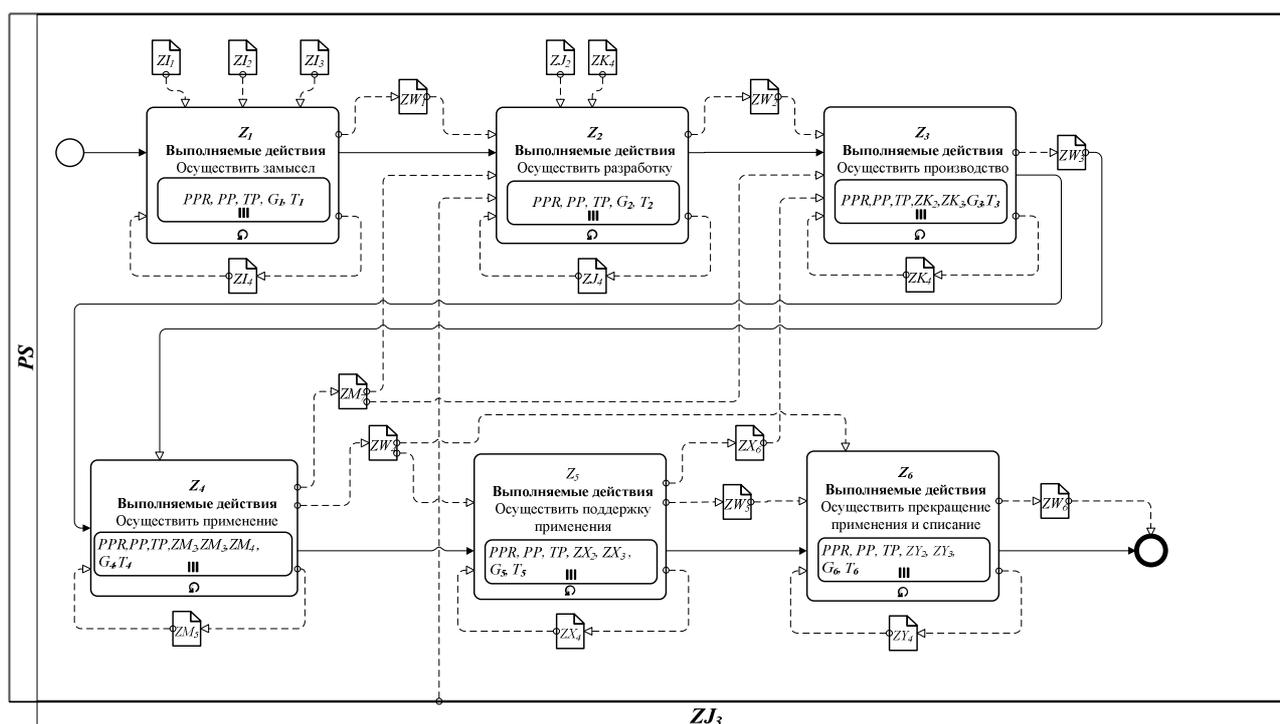


Рис. 9. Диаграмма типового описания процессов стадий жизненного цикла в нотации BPMN

Применение данного подхода позволяет сформулировать факторы, определяющие природу изменения эффективности и допустимости состояний сложной системы и свести массивы информации, описывающие состояние системы, к ограниченному числу обобщенных характеристик. Использование языка теории нечетких множеств в качестве метаязыка позволяет правильно отобразить основные свойства системы и упростить процесс ее анализа.

С точки зрения развития экономики страны человеческие ресурсы играют двудединую 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. Однако учетные процедуры проводятся одновременно различными ведомствами в соответствии с решаемыми ими задачами и, в связи с этим, охватывают различные сегменты населения. Всего в Российской Федерации в настоящее время существует

порядка 18 ведомственных баз данных, содержащих различное количество записей о физических лицах, численность которых также изменяется от базы к базе и составляет от нескольких миллионов до более 100 млн единиц учета. 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). Может быть определена лингвистическая переменная их определяющая, на основе значений которой определено нечеткое правило принадлежности данного объекта к тому или иному классу.

ЗАКЛЮЧЕНИЕ

Методика, предложенная для моделирования предметной области, реализует:

- Поэтапное выделение объектов предметной области.
- Установление связей между атрибутами.
- Сокращение нечеткости при каждой последующей стадии жизненного цикла.
- Методика моделирования может быть реализована:
 - Посредством SADT-методологии.
 - Использованием формальных языков, таких как BPMN/IDEF/UML.

СПИСОК ЛИТЕРАТУРЫ

1. **Айзерман М. А., Алесекров Ф. Т.** Выбор вариантов. Основы теории. М.: Наука, 1990. 240 с.
2. **Бениаминов Е. М.** Алгебраические методы в теории баз данных и представлении знаний // Научный мир. М., 2003. 184 с.
3. **Борисов А. Н., Крумберг О. А., Федоров И. П.** Принятие решений на основе нечетких моделей. Примеры использования. Зинатне, Рига : 1990. 180 с.
4. **Куликов Г. Г., Антонов В. В.** Метод формирования структуры хранилища данных для автоматизированной учетной системы на основе процессного анализа предметной области // Вестник УГАТУ. 2006. Т. 8, № 1(17). С. 60–67.
5. **Погонин В. А.** Модели диспетчерского управления роботами // Интернет-журнал «Информационные процессы и управление». 2006. № 1. С. 45–57.
6. **Поспелов Д. А.** Знания и шкалы в модели мира // Модели мира. М.: РАИИ, 1997. С. 69–84.
7. **Рыжов А. П.** Элементы теории нечетких множеств и ее приложений [Электронный ресурс] // Диалог-МГУ. (<http://intsys.msu.ru/staff/ryzhov/Fuzzy-SetsTheory&Applications.pdf>).
8. **Хомский Н.** Язык и проблема знания // Вестник МГУ. 1996. Вып. 6. С. 157–185.

ОБ АВТОРАХ

Антонов Вячеслав Викторович, доц. каф. автоматизированных систем управления. 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. инженер по автоматиз. машиностроения (УАИ, 1971). Д-р техн. наук по системн. анализу, автоматич. 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 автоматизированным системам обработки информации и управления (УГАТУ, 2010). Готовит дис. в обл. построения информационных систем.