

УДК 004.78:351.755

Г. Г. КУЛИКОВ, В. В. АНТОНОВ, А. А. САВИНА

ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА НА ОСНОВЕ ЗНАНИЙ ДЛЯ РЕГИСТРАЦИОННОГО УЧЕТА НАСЕЛЕНИЯ

Рассмотрены проблемы интеграции информации в распределенных системах для поддержки принятия управленческих решений и пути решения данных проблем. На основе представления бизнес-процессов в виде совокупности взаимодействующих вертикальных и горизонтальных семантически определенных и формализованных объектов предложен метод формирования структуры информационной системы. *Вертикальная автоматизация ; процессный анализ ; бизнес-процесс*

Проблема повышения эффективности и качества управления в социально-экономических системах является в настоящее время одной из наиболее обсуждаемых, на решение ее ориентированы ряд национальных проектов. Однако до настоящего времени остаются неисследованными многие теоретические аспекты, связанные с моделированием процессов автоматизации регистрационных регламентов и информационным сопровождением процессов интеграции учетной информации совместно с семантическими правилами их регламентирующими. Информация, находящаяся в распоряжении государства, должна обеспечивать реализацию прав и свобод граждан, а также принятие и реализацию государственными служащими управленческих решений. Государственный учет понимается как сбор и хранение органами государственной власти и местного самоуправления сведений, необходимых для выполнения государственных функций и оказания услуг гражданам и организациям. Примером учета может служить регистрационный учет населения. По сути, компьютерные программы, реализуя заложенные в них алгоритмы, представляют трактовку нормативной базы со стороны своих разработчиков. В связи с тем, что уже сейчас существует критичная зависимость государственных органов от используемого программного обеспечения, а также в виду необходимости корректировки программного обеспечения по мере изменения нормативной базы на фоне непрерывно увеличивающихся объемов информации, особую значимость приобретают исследования по применению CALS-технологий при разработках

распределенных учетных систем поддержки принятия управленческих решений для органов государственной власти и местного самоуправления.

1. ЗАДАЧИ ИССЛЕДОВАНИЯ

Корни CALS-стандартов уходят в обеспечение логистики. Изначально эта аббревиатура звучала как система автоматизированной поддержки логистики — Computer Aided Logistic Support, и только в последствии стала трактоваться как непрерывная поддержка жизненного цикла. CALS предписывает построение функциональных моделей на формальном языке IDEF, а основной метод, применяемый при функциональном моделировании, — метод декомпозиции. К настоящему времени CALS-технологии образуют самостоятельное направление в области информационных технологий. Существует множество объектов, информационные модели которых уже построены и стандартизованы (в соответствии с требованиями ISO), так, что при использовании этих объектов в качестве базовых при построении информационной системы можем ссылаться на стандарт STEP. Также можем использовать стандартные способы построения рабочей базы данных на основании информационной модели (Express-компилятор) с использованием стандартных методов доступа (SDAI) к Express-данным. Вся эта совокупность знаний объединена в понятии CALS. Анализ правил регистрационного учета населения позволяет выявить общие для многих регламентов и учетов характеристики, на основе которых может быть произведена структурная формализация с возмож-

ностью дальнейшей интеграции бизнес-процессов и процессов управления по семантическим правилам для формирования хранилища данных, что в итоге позволяет обеспечить разработку метода формирования распределенной учетной системы и построения модели хранилища данных для автоматизированной учетной системы на основе процессного анализа и классификации атрибутов предметной области по правилам нечеткой логики. На основе представления бизнес-процессов в виде совокупности взаимодействующих вертикальных и горизонтальных семантически определенных и формализованных объектов предложен метод формирования структуры информационной системы, отличающийся тем, что с целью направленного развития системы производится поэтапная автоматизация бизнес-процессов с постепенным «выращиванием» единой системы на основе мониторинга иерархических отношений классов атрибутов определяющих их бизнес-процессы. Для формирования четких измерений OLAP-кубов интегрированного хранилища данных к нечетким правилам перекрытия взаимодействующих объектов учета применены методы теории нечетких множеств.

2. ПОСТРОЕНИЕ МОДЕЛИ ИНФОРМАЦИОННОЙ СИСТЕМЫ

Одним из основных критериев качества информационной системы может служить адекватность организации объектов учета внутри информационной системы (согласно информационной модели) и реально присутствующей в виде множества данных, упорядоченных в отношениях между собой. Каждая учетная информационная система обеспечивает комплекс мероприятий по фиксации и актуализации объектов учета. Свойства каждого вида объекта, жизненный цикл имеют ряд индивидуальных особенностей. В целях поддержки информационной совместности необходимо обеспечить поддержку так называемой нейтральной модели данных, пригодной для представления разнообразных данных об объектах учета. В качестве такой модели в настоящее время выступает международный CALS-стандарт — ISO-10303 (STEP). Если представить учетную информационную систему в виде совокупности взаимодействующих бизнес-процессов (объяснив последние объектами учета) и классифицировать виды взаимодействия этих объектов учета, то появляется возможность рассмотрения информационной системы как ав-

томатизированной системы поддержки логистики. В рамках стандарта ISO-10303 возможна регламентация логической структуры создаваемого хранилища данных, номенклатуры информационных объектов и их атрибутов, а наличие методики расширения информационной модели данных позволяет адаптировать имеющуюся информационную модель в процессе выстраивания системы.

Анализ процессов реализации административных регламентов и связанные с этим выделение подпроцессов, операций и правил позволяют использовать их для идентификации функциональных модулей с дальнейшим представлением административного регламента в виде совокупности взаимодействующих вертикальных и горизонтальных семантически определенных и формализованных объектов. В первую очередь подлежат автоматизации государственные административные регламенты, рутинные и хорошо структурированные. Этот тип государственных процессов и регламентов очень похож на бизнес-процессы, характеризуется высокой степенью повторяемости (например, выдача паспорта).

На основании результатов приведенных анализов можно сформулировать основные принципы концепции эволюции существующих систем учета, предназначенных для повышения их эффективности:

1. Широкое применение электронных средств требует повышения степени формализации организационной структуры.

Основные информационные объекты, используемые в бизнес-процессах структурного подразделения сводятся в матрицу операций над ресурсами, для каждого объекта определяется структура его жизненного цикла, назначаются роли и их исполнители в соответствии с организационной структурой и штатным расписанием. Для инициализации и управления исполнением бизнес-процессов они дополняются необходимыми функциями управления (нормирование, планирование, регулирование, учет) и соответственно ролями и исполнителями этих функций.

Таким образом, за каждым бизнес-процессом с информационным объектом закрепляется и процесс управления:

$$MOR = \{IO\} \cap \{FIP\} \quad (1)$$

$$MFU = \{IO\} \cap \{FU\} \quad (2)$$

$$MR = \{Roles\} \cap \{FU\} \quad (3)$$

$$BP = \{IO, FIP, OS\} \quad (4)$$

$$PU = \{BP, FU, OS\}, \quad (5)$$

где IO — множество информационных объектов,

FIP — множество первичных информационных процессов,

FU — множество функции управления,

MOR — матрица операций над ресурсами,

MFU — матрица функций управления информационными объектами,

MR — матрица ролей для функции управления,

BP — бизнес-процесс,

PU — процесс управления,

OS — множество штатных единиц (штатная структура).

2. Повышение степени структурной и параметрической адекватности модели объекта учета путем регулярной организации анализа функциональной составляющей системы учета.

Моделью знаний объекта учета является хранилище данных, актуализируемое во времени и пространстве горизонтальными и вертикальными бизнес-процессами. Таким образом может быть реализован адекватный процессный мониторинг за счет изменения как самой структуры, так и хранилища информации, поддерживающий интеграцию репозитория моделей и хранилища данных на основе правил нечеткой логики для аналитического анализа с применением OLAP-технологий.

Модель, охватывающая информационную систему, может быть представлена в виде метабазы, в которой содержится информация по каждому виду объекта учета, в нашем случае о каждом экземпляре независимого бизнес-процесса (обозначим как множество G).

С другой стороны информационная система представима в виде функциональной системы — т. е. в виде множества функций (обозначим как множество F).

Введем обозначения:

Q_G — множество свойств, определяемых отношениями между элементами множества G .

Q_F — множество свойств, определяемых отношениями между элементами множества F .

Q_{FG} — множество свойств, определяемых связями между элементами множеств F и G .

Тогда взаимосвязи внутри множеств F и G могут быть определены отношениями на декартовом произведении

$$Q_F \times F = \left\{ \begin{array}{l} z_i^F = (q_i^F, f_i) : q_i^F \in Q_F, \\ f_i \in F, i = 1, \dots, n \end{array} \right\} \quad (6)$$

и

$$Q_G \times G = \left\{ \begin{array}{l} z_i^G = (q_i^G, g_i) : q_i^G \in Q_G, \\ g_i \in G, i = 1, \dots, n \end{array} \right\} \quad (7)$$

соответственно.

Взаимосвязь между ними может быть определена отношением на декартовом произведении

$$Q_{FG} \times G \times F = \left\{ \begin{array}{l} z_i^{FG} = (q_i^{FG}, g_i, f_i) \\ : q_i^{FG} \in Q_{FG}, \\ g_i \in G, f_i \in F, i = 1, \dots, n \end{array} \right\}. \quad (8)$$

Принадлежность элемента z_i^{FG} данному отношению может интерпретироваться следующим образом: «в элементе g_i информационной системы содержится информация по свойству q_i^{FG} функциональной части информационной системы f_i ».

Поиск информации, соответствующей конкретному элементу f_i в g_i , сводится к определению отношения $R \subseteq G \times F$. Таким образом, о любой паре $(g_i, f_i) \in R : g_i \in G, f_i \in F, i = 1, \dots, n$ можно сказать, что f_i является релевантным g_i , и решение задачи определения релевантности элементов множеств G и F , сводится к определению отношения $R \subseteq G \times F$. При этом для любых $g_i \in G, f_i \in F, g_j \in G, f_j \in F, i, j = 1, \dots, n$, если выполняются условия $f_i \subseteq f_j$ и $g_i \subseteq g_j$, то есть все элементы g_i содержатся g_j и все элементы f_i содержатся в f_j и $(g_i, f_i) \in R$, то условие $(g_j, f_j) \in R$ также выполняется. За исключением крайнего случая, когда отношение R есть само декартово произведение $G \times F$, 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 R можно поставить в соответствие логическое выражение (предикат) Q_{FG} , зависящее от определенного числа параметров (n -местный предикат) и определяющее, будет ли кортеж (g_j, f_j) принадлежать отношению R . Таким образом, принадлежность кортежа отношению эквивалентна истинности предиката $(g_j, f_j) \in R \Leftrightarrow \{Q_{FG}\} = \{G, F, R\}$.

В качестве схем информационных объектов рассматриваемой модели определим независимые бизнес-процессы. Тогда правила взаимодействия экземпляров бизнес-процессов могут оцениваться двумя типами свойств:

те, которые можно непосредственно измерить, и те, которые являются качественными и требуют попарного сравнения объектов, обладающих оцениваемым свойством, чтобы определить их место по отношению к рассматриваемому понятию.

Учитывая построение нашей информационной системы в виде объектов, каждый независимый бизнес-процесс можем рассматривать как отдельную (обособленную) часть информационной системы. 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 вышесказанное, 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 данного уровня иерархии;
- по вертикали — бизнес-процессы, которые взаимодействуют с вышестоящей иерархией по семантическим правилам согласно функциональному назначению.

Типовая организационно-функциональная модель рассматриваемого уровня иерархии с разнесением на условные семантические блоки приведена на рис. 1, где *TD* — территориальное подразделение, *FD* — функциональное.

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

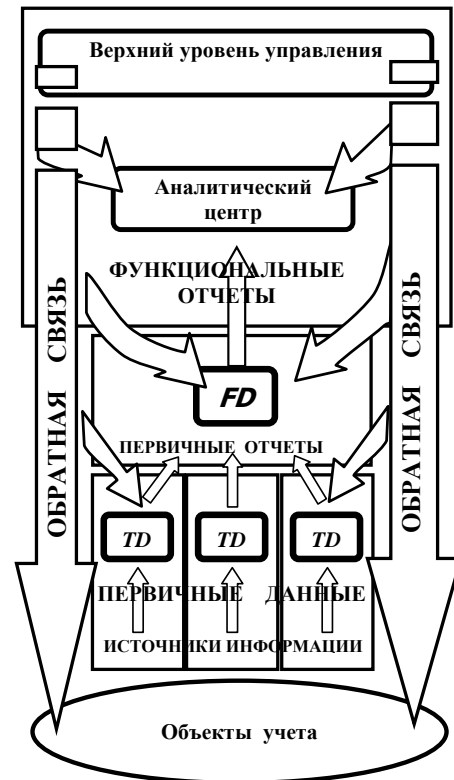


Рис. 1. Структура организации на основе формального разнесения по семантическим правилам на объекты

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.

Анализируя опыт внедрения систем, можно сделать вывод о том, что в ряде случаев острота вышеуказанной проблемы может быть снижена за счет отказа от «революционного» пути внедрения в пользу так называемого «эволюционного». Рассмотрим возможность совмещения поэтапной автоматизации с процессным подходом, т. е. первоначально проведем обследование, целью которого будет выявление независимых бизнес-процессов организации. Далее рассмотрим вопрос автоматизации выбранного бизнес-процесса

на всех уровнях иерархии. Фрагмент модели автоматизации выбранного бизнес-процесса приведен на рис. 2.

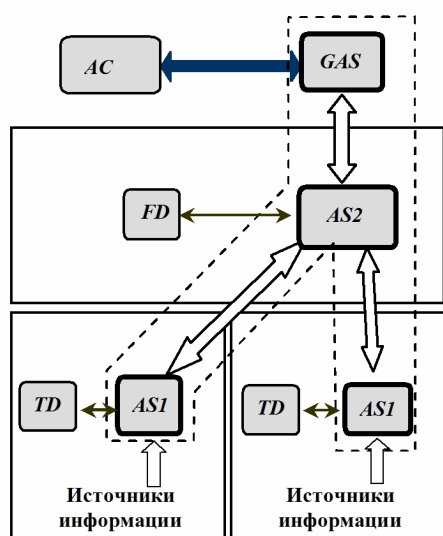


Рис. 2. Фрагмент модели автоматизации выбранного бизнес-процесса

Введем следующие обозначения:

AC — аналитический центр;

GAS — главная автоматизированная система;

FD — функциональное подразделение;

TD — территориальное подразделение;

AS2 — автоматизированная система, обеспечивающая функционирование выбранного бизнес-процесса;

AS1 — автоматизированная система, обеспечивающая функционирование выбранного бизнес-процесса территориального подразделения.

При таком подходе информация становится доступной руководителю для анализа сразу же после ее ввода исполнителем. При этом полнота информации, доступной на всех уровнях иерархии, растет по мере увеличения числа автоматизированных бизнес-процессов. Очередность автоматизации бизнес-процессов определяется потребностями управляющего звена верхнего уровня. Отдельные компоненты системы (выделенные объекты) могут дорабатываться и адаптироваться, оставаясь объединенными в единую систему, которая будет одновременно и устойчивой и гибкой. Каждый компонент системы может развиваться самостоятельно, не нарушая системные связи, и, следовательно, перенастройка отдельных компонентов системы не повлияет на работоспособность системы в целом.

Затем с учетом полученных результатов производится автоматизация следующего бизнес-процесса, чем обеспечивается возможность поэтапной автоматизации бизнес-процессов организации с постепенным «выращиванием» единой системы, охватывающей все бизнес-процессы.

Очевидно, что при поэтапной автоматизации отдельных бизнес-процессов, значительно снижается нагрузка на персонал и минимизируются неизбежные организационные и психологические проблемы внедрения, обеспечивается повышение степени структурной и параметрической адекватности модели объекта учета путем регулярной организации анализа функциональной составляющей системы учета. Моделью знаний объекта учета является хранилище данных, актуализируемое во времени и пространстве горизонтальными и вертикальными бизнес-процессами. Таким образом может быть реализован адекватный процессный мониторинг за счет изменения как самой структуры, так и хранилища информации. Управление любой сложной социально-экономической системой невозможно без обратной связи, которая заключается в отслеживании и анализе данных, отражающих состояние этой системы и ситуацию вокруг нее. Постоянная доступность актуальной информации дает возможность оценить текущее положение дел, а обзор изменения конкретных характеристик во времени позволяет обнаружить тенденции развития. Такое управление основано на знании и потому наиболее эффективно [2].

В качестве схем информационных объектов рассматриваемой модели определены независимые бизнес-процессы, которые характеризуются значениями своих атрибутов. Поэтому каждый рассматриваемый бизнес-процесс может быть определен как множество упорядоченных пар:

$$x = \{ \langle a_i, d_i \rangle \mid i \in I \}, \quad (9)$$

где I — множество индексов, a — имя атрибута, d — значение атрибута.

Каждый из атрибутов может принадлежать некоторому классу:

- идентификационные (ключевые) атрибуты (обозначим A^1);
- функциональные атрибуты (обозначим A^2);
- неформализованные атрибуты (задающиеся вербальным описанием, имеющие ка-

ественные характеристики) (обозначим A^3).

$$x = \left\{ \begin{array}{l} \langle a_1^1, d_1^1 \rangle, \dots, \langle a_k^1, d_k^1 \rangle, \\ \langle a_1^2, d_1^2 \rangle, \dots, \langle a_l^2, d_l^2 \rangle, \\ \langle a_1^3, d_1^3 \rangle, \dots, \langle a_m^3, d_m^3 \rangle \\ : k, l, m \in I \mid x = k + l + m \end{array} \right\} \quad (10)$$

Введем обозначения:

$\varphi : A \rightarrow D$ — отображение, ставящее в соответствие каждому атрибуту множество его возможных значений ($\varphi : A^1 \rightarrow D^1$ — для каждого идентификационного атрибута, $\varphi : A^2 \rightarrow D^2$ — для каждого функционального атрибута, $\varphi : A^3 \rightarrow D^3$ — для каждого неформализованного атрибута).

Множества тех или иных классов атрибутов могут частично, полностью или совсем не пересекаться в разных объектах учета. По отношению двух бизнес-процессов друг к другу при интеграции их можно классифицировать следующим образом:

- Однородные бизнес-процессы — это бизнес-процессы, множества идентификационных и функциональных классов атрибутов которых одинаковы:

$$\{A^1\}^1 = \{A^1\}^2 \text{ и } \{A^2\}^1 = \{A^2\}^2. \quad (11)$$

- Пересекающиеся однородные бизнес-процессы — это бизнес-процессы, множества идентификационных классов атрибутов которых одинаковы, а множества функциональных классов атрибутов пересекаются

$$\begin{aligned} \{A^1\}^1 = \{A^1\}^2 \text{ и } \{A^2\}^1 \neq \{A^2\}^2 \text{ и} \\ \{A^2\}^1 \cap \{A^2\}^2 \neq \emptyset. \end{aligned} \quad (12)$$

- Аддитивные бизнес-процессы — это бизнес-процессы, множества идентификационных классов атрибутов которых одинаковы, а множества функциональных классов атрибутов различаются:

$$\{A^1\}^1 = \{A^1\}^2 \text{ и } \{A^2\}^1 \neq \{A^2\}^2. \quad (13)$$

- Пересекающиеся бизнес-процессы — это бизнес-процессы, множества идентификационных классов атрибутов которых пересекаются, а множества функциональных классов атрибутов одинаковы:

$$\{A^2\}^1 \cap \{A^2\}^2 = \emptyset \text{ и } \{A^2\}^1 = \{A^2\}^2. \quad (14)$$

- Пересекающиеся бизнес-процессы с элементами аддитивности — это бизнес-процессы, множества идентификационных классов атрибутов которых пересекаются, а множества функциональных классов атрибутов

различаются:

$$\{A^2\}^1 \cap \{A^2\}^2 = \emptyset \text{ и } \{A^2\}^1 \neq \{A^2\}^2. \quad (15)$$

- Разнородные пересекающиеся бизнес-процессы — это бизнес-процессы, множества идентификационных классов атрибутов которых не пересекаются и множества функциональных классов атрибутов одного бизнес-процесса пересекаются со множеством идентификационных классов атрибутов другого бизнес-процесса, что позволяет провести интеграцию

$$\begin{aligned} \{A^1\}^1 \cap \{A^1\}^2 = \emptyset \\ (\{A^1\}^1 \cap \{A^2\}^2) \cup (\{A^2\}^1 \cap \{A^1\}^2) \neq \emptyset. \end{aligned} \quad (16)$$

- Разнородные непересекающиеся бизнес-процессы — это бизнес-процессы, множества идентификационных классов атрибутов которых не пересекаются и множества функциональных классов атрибутов одного бизнес-процесса не пересекаются со множеством идентификационных классов атрибутов другого бизнес-процесса

$$\begin{aligned} \{A^1\}^1 \cap \{A^1\}^2 = \emptyset \\ (\{A^1\}^1 \cap \{A^2\}^2) \cup (\{A^2\}^1 \cap \{A^1\}^2) = \emptyset. \end{aligned} \quad (17)$$

Учитывая построение нашей информационной системы в виде объектов, каждый независимый бизнес-процесс будем рассматривать как отдельную (обособленную) часть информационной системы.

Введем обозначения:

$T = \{t_i : i \in I\}$ — множество типов независимых бизнес-процессов (типов объектов);

$\delta : T \rightarrow 2^A$ — отображение, задающее для каждого типа бизнес-процесса (типа объекта) множество его атрибутов;

$U = \{x_i : i \in I\}$ — множество информационных объектов (независимых бизнес-процессов);

$\gamma : U \rightarrow T$ — отображение, ставящее в соответствие объекту его тип.

Каждый выделенный объект будет характеризоваться своей «информационной схемой» [4–6, 12], определяемой кортежем

$$S = \langle \{A^1, A^2, A^3\}, \{D^1, D^2, D^3\}, \varphi, T, \delta \rangle. \quad (18)$$

Информационная система объекта, построенная по этой схеме, будет определяться кортежем

$$U^S = \langle \{A^1, A^2, A^3\}, \{D^1, D^2, D^3\}, \varphi, T, \delta, U, \gamma \rangle. \quad (19)$$

Определим множество всех информационных систем объектов

$$U^{\bar{S}} = \left\{ U_i^{S_j} : i, j \in I \right\}, \text{ где } \bar{S} = \{S_j : j \in I\}. \quad (20)$$

Таким образом, под информационной системой выбранного объекта понимается «информационная схема объекта», описывающая характеристики входящих в нее бизнес-процессов и множество самих бизнес-процессов [4, 9]. Рассматривая, что изменится при автоматизации второго (третьего и т. д.) бизнес-процессов получаем, что в зависимости от типа подлежащих интеграции бизнес-процессов будет применяться один из вариантов приведенного выше взаимодействия бизнес-процессов, и обуславливаться изменением числа атрибутов множества идентификационных и функциональных классов [6].

В зависимости от типа подлежащих интеграции бизнес-процессов будет применяться один из вариантов, представленный уравнениями (11)–(17), и обуславливаться изменением числа атрибутов множества идентификационных и функциональных классов:

- для однородных бизнес-процессов и пересекающихся однородных бизнес-процессов интеграция производится без изменения числа атрибутов множества идентификационных и функциональных классов;
- для аддитивных бизнес-процессов интеграция производится без изменения числа атрибутов множества идентификационных классов. Множество функциональных классов будет определяться формулой

$$\{A_1^2\}^1 = \{A^2\}^1 \cup \{A^2\}^2, \quad (21)$$

где A_1 атрибут интегрированного бизнес-процесса. Число атрибутов множества функциональных классов будет увеличено на количество несовпадающих атрибутов.

Для пересекающихся бизнес-процессов интеграция производится по совпадающим атрибутам идентификационных классов, при этом множество идентификационных классов будет определяться формулой

$$\{A_1^1\}^1 = \{A^1\}^1 \cup \{A^1\}^2. \quad (22)$$

Число атрибутов множества идентификационных классов будет увеличено на количество несовпадающих атрибутов. Интеграция производится без изменения числа атрибутов множества функциональных классов.

Для пересекающихся бизнес-процессов с элементами аддитивности интеграция производится по совпадающим атрибутам идентификационных классов, при этом множество идентификационных классов будет определяться формулой

$$\{A_1^1\}^1 = \{A^1\}^1 \cup \{A^1\}^2. \quad (23)$$

Число атрибутов множества идентификационных классов будет увеличено на количество несовпадающих атрибутов. Множество функциональных классов будет определяться формулой

$$\{A_1^2\}^1 = \{A^2\}^1 \cup \{A^2\}^2. \quad (24)$$

Число атрибутов множества функциональных классов будет увеличено на количество несовпадающих атрибутов.

Для разнородных пересекающихся бизнес-процессов интеграция производится по совпадающим атрибутам идентификационных классов одного бизнес-процесса со значениями функциональных атрибутов другого бизнес-процесса. При этом множество идентификационных классов будет определяться формулой

$$\{A_1^1\}^1 = \{A^1\}^1 \cup (\{A^1\}^1 \cap \{A^2\}^2). \quad (25)$$

Число атрибутов множества идентификационных классов будет увеличено на количество совпадающих атрибутов идентификационных классов одного бизнес-процесса со значениями функциональных атрибутов другого бизнес-процесса. Интеграция производится без изменения числа атрибутов множества функциональных классов.

Для разнородных непересекающихся бизнес-процессов интеграция производится в виде совместного хранения полностью независимых учетов.

Отношения между взаимодействующими бизнес-процессами оцениваются двумя типами свойств: теми, которые можно непосредственно измерить, и теми, которые являются качественными и требуют попарного сравнения объектов, обладающих оцениваемым свойством, чтобы определить их ме-

сто по отношению к рассматриваемому понятию. Вопрос об интеграции на базе неформализованных атрибутов и непротиворечивости при этом информационной системы верхнего уровня сводится к рассмотрению зависимостей между типами входящих нее элементов. 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 шкалам [5, 6]. Так, практически в любой области можно получить множество шкал оценок, используя следующую процедуру:

- для оценки объекта определить список свойств;
- найти в этом списке полярные свойства и сформировать полярную шкалу;
- для каждой пары полюсов оценить, в какой степени введенное понятие обладает положительным свойством.

Каждому типу взаимодействующих объектов поставим в соответствие множество семантических зависимостей с другими типами объектов $C_i = \{c_{ij} : j \in I\}$. Каждый тип взаимодействующих объектов T_i , элементы которого удовлетворяют свойству C_i , определяется как множество

$$T_i = \left\{ \left\langle A^1, D^1 \right\rangle, \left\langle A^2, D^2 \right\rangle, \left\langle A^3, \mu_{C_{ij}}(A^3) \right\rangle : j \in I \right\}, \quad (26)$$

где A — множество атрибутов, D — множество значений атрибутов.

Определенные таким образом правила перекрытия (пересечения) взаимодействующих объектов учета и полученные на пересечении множеств новые множества определяют, в конечном счете, возможность использования OLAP-технологии на всем информационном пространстве интегрированного хранилища данных [4, 6]. Учитывая, что сами семантические ограничения (правила) могут быть заданы набором признаков в виде попарных сравнений, не для всех из которых существует однозначность ответа, для их анализа наиболее целесообразно использовать аппарат нечеткой логики. В нашем случае непротиворечивость рассматриваемой системы будет определяться отсутствием правил, имеющих сходные посылки и различные или взаимоисключающие следствия.

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 интеграции w_i , в каждый детерминированный момент времени будет являться элементом множества интеграционных процессов W . Тогда динамика информационной системы при переходе из одного состояния в другое будет описываться функцией F отображения $Q \times W \rightarrow Q$

$$F : Q \times W \rightarrow Q, \quad (27)$$

т. е. $f(q_i, w_i)$ есть последующее состояние информационной системы после выполнения интеграционного процесса w_i двух бизнес-процессов

$$q_{i+1} = f(q_i, w_i), \quad i = 1, \dots, n. \quad (28)$$

Начальное состояние q_0 будет определяться первым из автоматизируемых бизнес-процессов. Так как число шагов интеграции будет на единицу меньше числа бизнес-процессов, то число состояний информационной системы будет равно числу автоматизируемых бизнес-процессов. При этом на правила интеграции для каждого случая накладываются семантические ограничения — правила взаимодействия с другими типами объектов.

В общем случае можем сказать, что при каждом шаге построения информационной системы может быть описана так называемая нечеткая цель [7, 8] — нечеткое множество. Присутствуют правила-ограничения интеграции $C_i = \{c_{ij} : j \in I\}$, 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. Таким образом, изменение последовательности интеграции бизнес-процессов и последовательности ограничивающих правил не сказывается на цели.

Результатом взаимодействия двух бизнес-процессов является новый бизнес-процесс, который в свою очередь при очередном взаимодействии порождает новый бизнес-процесс. Эти «промежуточные» бизнес-процессы могут быть представлены в виде вирту-

альных организаций (подразделений), которые характеризуются отсутствием фиксированной организационной и территориальной структуры, в которых процесс создания учетной информации (данных, метаданных и правил их получения), может быть распределен во времени и пространстве между многими объектами учета. При этом ключевая идея создаваемой информационной системы совпадает с базовой идеологией систем, разрабатываемых на основе CALS и заключается в создании единой сложной модели получаемого бизнес-процесса.

Для каждого типа взаимодействующих объектов T_i степень непротиворечивости j -го и k -го правил можно задавать величиной

$$C_{jk}^{T_i} = \left\{ \left| \begin{array}{l} \max(\mu_{C_{ij}}(A_i^3), \mu_{C_{ik}}(A_i^3)) - \\ \max(\mu_{C_{ij}}(D_i^3), \mu_{C_{ik}}(D_i^3)) \end{array} \right| \right\}, \quad j, k \in I, j \neq k \quad (29)$$

3. СТРУКТУРА И ОСОБЕННОСТИ АППАРАТНО-ПРОГРАММНОЙ РЕАЛИЗАЦИИ АНАЛИТИЧЕСКОЙ СИСТЕМЫ

Одним из важнейших моментов является построение хранилищ данных с использованием OLAP-кубов для достижения максимальной производительности. Хранилище данных выступает в роли совокупности данных, объединенных из различных источников, структурированных и оптимизированных для доступа к ним при помощи OLAP-технологии. При этом обеспечивается и консолидация данных из различных источников, и разгрузка основной системы оперативной обработки [11]. Отличительной чертой создаваемого хранилища данных учетной системы является четкая предметная ориентация, что позволяет существенно ускорить доступ к данным за счет предварительной обработки данных и дополнительной реструктуризации в момент загрузки, позволяет также хранить только данные, необходимые для средств анализа, что существенно сокращает затраты на носители информации, повышает безопасность и скорость доступа. Первым этапом построения хранилища данных является логическое проектирование, состоящее в выявлении предметных областей, которые необходимо отразить в хранилище, и в выработке набора согласованных инструкций. На этапе физического проектирования структуры хранилища данных подготавливаются и запускаются в работу системы, обеспечивающие регулярное пополнение хранили-

ща данными из оперативных систем. Таким образом, хранилище данных обеспечивает интеграцию оперативных данных с помощью согласованных правил именования, измерений, физических атрибутов и семантики. Несмотря на все различия хранилищ данных, технологии процесса создания хранилищ имеют общие признаки. Для сложных систем характерно то, что управлять ими приходится, как правило, в условиях неполной информации, незнания закономерностей функционирования и постоянного изменения внешних факторов. За счет обратной связи, основанной на откликах пользователей, выявленных в ходе анализа различных закономерностей и аспектов функционирования, архитектура хранилища со временем претерпевает изменения, процесс создания хранилища становится итеративным [10, 11]. Таким образом, применение технологий OLAP для анализа в данном случае позволяет итеративно адаптировать хранилище под конкретные нужды и задачи организации.

Учитывая наличие нескольких центров актуализации (практически в каждом выделенном нами блоке), жестко регламентированных сроков и видов регламентной отчетности (по исходящей снизу вверх), наиболее целесообразно обеспечить одновременное хранение в рамках интегрированной информационной системы верхнего уровня детализированных данных, метаданных и агрегированных в виде аналитических таблиц данных. Учитывая, что на разных уровнях (слоях) и в каждом выделенном нами блоке может действовать свой собственный регламент отчетности, необходимо использование многоуровневой архитектуры хранилища данных. Решение данной проблемы может быть получено путем введения слоя навигационных метаданных между физическим хранилищем и программными приложениями его обработки, который к тому же обеспечивал бы оперативную трансляцию реляционных данных. Наиболее целесообразно введение в архитектуру хранилища промежуточного склада данных и использование платформенного разделения учета и отчетности, при котором учетная система оптимизирована для отражения бизнес-процессов, а отчетная является надстройкой над данными учетной системы и предназначена для быстрого создания и выпуска разнообразных отчетов.

Таким образом, построение хранилища данных выполняется в трехуровневой архитектуре.

На первом уровне расположены источники данных – внутренние регистрирующие, учетные и справочные системы.

На втором уровне содержится:

- центральное хранилище, куда стекается информация с первого уровня;
- промежуточный склад данных, который не содержит исторических данных и является источником аналитической информации для создания OLAP-кубов и витрин данных по разным направлениям регламентной отчетности.

На третьем уровне расположен набор предметно-ориентированных витрин данных, источником информации для которых является центральное хранилище данных и витрин данных по разным направлениям регламентной отчетности, источником информации для которых является промежуточный склад данных.

Средствами доступа к многомерным кубам данных обычно служат конструкторы запросов, позволяющие получать отчеты в виде таблиц путем фиксирования необходимых меток для получения срезов данных. При этом процедура формирования запроса осуществляется независимо от того, есть ли в гиперкубе запрашиваемые данные или нет, и результатом выборки данных может стать пустое значение.

Практика разработки и внедрения территориально-распределенных систем сбора учетных данных показала, что сбор первичных данных осуществляется поэтапно в течение продолжительного времени. Поэтому создаваемые многомерные кубы данных (гиперкубы) имеют низкую плотность заполнения данными и являются разреженными. Зависимость затрачиваемого времени от числа меток в измерении схематично приведена на рис. 3.

Можем разбить данный график на конечное число интервалов, в которых интервал изменения затрат времени одинаков.

Введем обозначения: k – число измерений гиперкуба данных, m_i – число меток в i -м измерении, t_i – время, затраченное на фиксацию метки i -го измерения, f_j – среднее значение затрачиваемого времени при фиксации метки в j -м интервале (разбитие по количеству меток в измерении).

Зависимость f_j между собой может быть выражена формулой

$$f_k = f_j + \Delta f \times (k - j), \quad (30)$$

где Δf – разница между соседними значениями f_j .

Очевидно, что

$$t_i = F(f_1, \dots, f_l, m_i) = \begin{cases} f_1, & m_i \in [r_1, r_2) \\ f_2, & m_i \in [r_2, r_3) \\ \dots \\ f_l, & m_i \in [r_{l-1}, r_l) \end{cases}, \quad (31)$$

где r_n – границы интервалов, l – число интервалов.

Общее количество затрачиваемого времени на фиксирование меток всех измерений может быть выражено формулой

$$t_{\text{общ}} = \sum_{i=1}^k t_i = \sum_{i=1}^k F(f_1, \dots, f_l, m_i). \quad (32)$$

Таким образом, $t_{\text{общ}}$ может быть представлено произвольным набором f_j , общим количеством k штук. Формула (34) может быть приведена к виду

$$t_{\text{общ}} = k \times f_1 + k \times \Delta l \times \Delta f, \quad \Delta l \in [0, l - 1]. \quad (33)$$

Учитывая, что Δl линейно зависит от m_i , полученная зависимость $t_{\text{общ}}$ может быть интерпретирована следующим образом: функция, отражающая время, затраченное на фиксацию меток всех измерений, растет быстрее увеличения числа измерений на величину, пропорциональную произведению числа измерений на число меток.

Действенным способом повышения эффективности в этом случае является оптимизация формирования пользовательского запроса к разреженному гиперкубу, по сути – создание алгоритма, позволяющего избежать получения пустой выборки на этапе формирования запроса.

Все данные в хранилище данных делятся на три основных категории:

- метаданные;
- детальные данные;
- агрегированные данные.

Наличие метаданных позволяет осуществлять быструю и удобную навигацию по различным уровням данных, позволяет наглядно представлять реализованную структуру хранилища данных и т. д. Проведенные исследования показали, что большинство конечных пользователей предпочитают работать с агрегированными показателями.

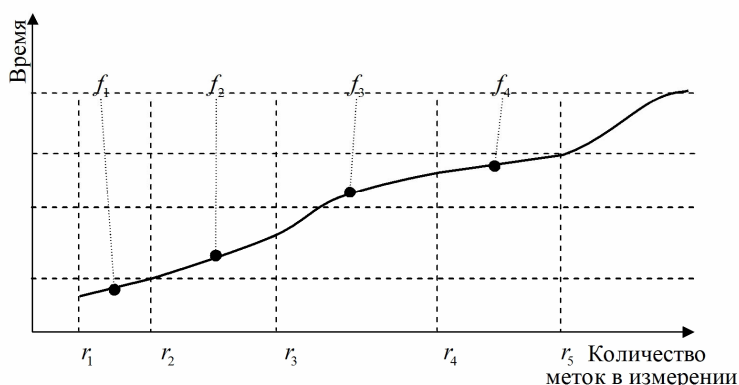


Рис. 3. График зависимости времени фиксации метки от числа меток в измерении

В структуре хранилища данных предусматривается возможность получения агрегированной информации с последующей навигацией по всем уровням агрегирования.

Концептуальная модель «хранилища данных» приведена на рис. 4.

Таким образом, описываемый вариант хранилища данных создается в виде самостоятельной системы для «верхнего слоя» и в качестве надстройки над множеством существующих систем обработки данных «среднего» и «нижнего» слоев, где могут функционировать и проблемно-ориентированные (ведомственные) реляционные базы данных и унаследованные системы.

Базируясь на предложенной классификации атрибутов и методе формирования структуры информационной системы в виде совокупности взаимодействующих вертикальных и горизонтальных семантически определенных и формализованных объектов, связанных друг с другом иерархическими отношениями классов атрибутов определяющих их бизнес-процессов, наиболее целесообразно использовать в качестве измерений при построении куба множеств, полученных согласно правил перекрытия взаимодействующих объектов учета. Таблица фактов при этом содержит целочисленные колонки, дающие числовую характеристику каждого, определенного таким образом измерения, и несколько целочисленных колонок — ключей для доступа к таблицам измерений, которые их расшифровывают.

Для каждого измерения составляем список уникальных значений из элементов, хранящихся в столбцах и производим предварительное агрегирование фактов для записей, имеющих одинаковые значения размерностей. Используя промежуточные таблицы (так называемые кросс-таблицы) можем свя-

зать элементы разных таблиц между собой, для чего каждой записи в таблицах измерений поставим в соответствие список, элементами которого будут номера фактов, при формировании которых использовались эти измерения. Для фактов соответственно каждой записи поставим в соответствие значения координат, по которым она расположена в гиперкубе. Измерения имеют иерархическую структуру, состоящую из одного или нескольких уровней, на основании которой осуществляются операции свертки или детализации.

Для каждого свободного бизнес-процесса может создаваться отдельная OLAP-таблица, логическое связывание которых и построение OLAP-таблиц следующего уровня возможно уже только на уровне агрегированных аналитических показателей, приведенных к сопоставимой форме.

Для построения важны уникальные значения, имеющиеся в полях измерений. Для построения срезов гиперкуба нам необходимы следующие возможности — определение координат (фактически значения измерений) для записей таблицы, а также определение записей, имеющих конкретные координаты (значения измерений). Действия разбиты на два этапа — согласно приведенной классификации атрибутов. Для идентификационных и функциональных атрибутов можем сразу составить таблицу уникальных значений, на основании которой будет производиться построение измерения. В зависимости от типа подлежащих интеграции бизнес-процессов, применяя один из вариантов представленный уравнениями (12)–(17), определяем правило интеграции в соответствии с формулами (21)–(25). Для неформализованных атрибутов, на основе установленных семантических зависимостей между типами объектов учета определяем правило интеграции.

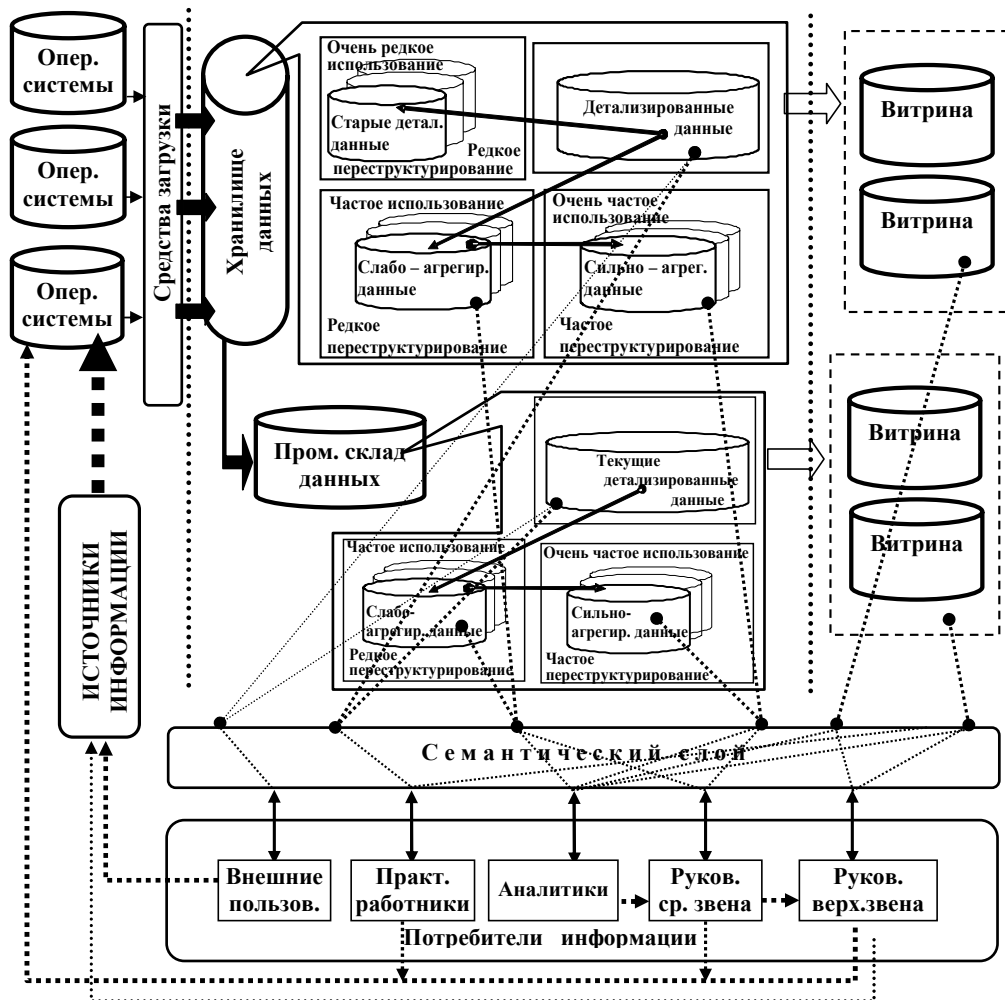


Рис. 4. Концептуальная модель хранилища данных

Далее определяем ключевые слова (фразы), составляем таблицу совокупности вариантов их комбинаций, удовлетворяющих семантическому правилу (по функции принадлежности), и, как и в случае с формализованными атрибутами, строим таблицу уникальных значений. То есть вместо одной таблицы мы получили аналог нормализованной базы данных. Нас интересуют только координаты в нашем гиперкубе, поэтому определим координаты для значений измерений. Самым простым будет перенумеровать значения элементов.

ЗАКЛЮЧЕНИЕ

На основании результатов проведенных анализов предложено решение задачи формирования структуры информационной системы новым способом представления бизнес-процессов в виде совокупности взаимодействующих вертикальных и горизонтальных семантически определенных и формализо-

ванных объектов применительно к новым условиям использования иерархических отношений классов атрибутов определяющих их бизнес-процессы. Разработанная модель, метод и алгоритмы поэтапной автоматизации позволяют преобразовать действующие бизнес-процессы в сквозные для выбранного семантического блока, а все блоки соединить между собой в единую функциональную систему.

Использование предложенной классификации атрибутов позволяет повысить формализованность правил взаимодействия объектов учета, а задачу построения информационной аналитической системы свести к формальному алгоритму, что открывает возможность поэтапного целенаправленного конструирования автоматизированной системы.

Представленный метод позволяет применить его к широкому спектру задач автоматизации процессов предоставления государственных услуг населению и учета выполненных работ.

СПИСОК ЛИТЕРАТУРЫ

1. **Абаев, Л. Ч.** Выбор вариантов в нечеткой среде: бинарные отношения и нечеткая декомпозиция / Л. Ч. Абаев [Электронный ресурс] (<http://fuzzy.kstu.ru/fullproc.htm>).
2. **Анфилатов, В. С.** Системный анализ в управлении / В. С. Анфилатов, А. А. Емельянов, А. А. Кукушкин. М. : Финансы и статистика, 2006. 367 с.
3. **Борисов, А. Н.** Принятие решений на основе нечетких моделей. Примеры использования / А. Н. Борисов, О. А. Крумберг, И. П. Федоров. Рига : Зинатне, 1990. 184 с.
4. **Воробьева, М. С.** Построение модели интеграции данных в информационно-управляющих системах / М. С. Воробьева // Модернизация образования в условиях глобализации : круглый стол «Образование через науку и инновации», 14–15 сент. 2005. Тюмень : ТюмГУ, 2005. С. 26–28.
5. **Куликов, Г. Г.** Определение структуры многоуровневой территориально-распределенной учетной системы на основе формальной интеграции бизнес-процессов объектов по семантическим правилам / Г. Г. Куликов, В. В. Антонов // Управление экономикой: методы, модели, технологии : матер. 6-й всерос. науч. конф. Уфа, 2006. Т. 1. С. 228–233.
6. **Куликов, Г. Г.** Метод формирования структуры хранилища данных для автоматизированной учетной системы на основе процессного анализа предметной области / Г. Г. Куликов, В. В. Антонов // Вестник УГАТУ. 2006. Т. 8. С. 60–67.
7. **Пивкин, В. Я.** Нечеткие множества в системах управления / Е. П. Пивкин, Е. П. Бакулин, Д. И. Кореньков [Электронный ресурс]. (http://idisys.iae.nsk.su/fuzzy_book/content.html).
8. **Паклин, Н. В.** Нечеткая логика — математические основы / Н. В. Паклин [Электронный ресурс]. (<http://basegroup.ru/fuzzy-logic/math.htm>).
9. **Цаленко, М. Ш.** Моделирование семантики в базах данных / М. Ш. Цаленко. М. : Наука, 1989. 288 с.
10. **Чугреев, В. Л.** Модель структурного представления текстовой информации и метод ее тематического анализа на основе частотно-контекстной классификации : дис. ... канд. техн. наук / В. Л. Чугреев. С.-Петербург. гос. электротехн. ун-т «ЛЭТИ», 2003.
11. **WWW.** Совместное использование учетных систем и технологии OLAP [Электронный ресурс]. (citforum.oldbank.com/database/articles/olap_oltp.html).

ОБ АВТОРАХ



Куликов Геннадий Григорьевич, проф., зав. каф. АСУ. Дипл. инж. 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 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.



Савина Анна Александровна, инж.-прогр. ИЦ МВД по РБ. Дипл. экон. по инф. системам в экономике (УГАТУ, 2004). Иссл. в обл. автоматиз. информ. систем.