

УДК 004.62; 025.4.03

# Управление тематическими данными в больших и сверхбольших хранилищах: механизмы, модели, программное обеспечение (состояние, задачи, решения)

В. А. Васенин

Московский государственный университет им. М. В. Ломоносова, 119992, Москва, Россия

Научно-исследовательский институт механики Московского государственного университета, 119992, Москва, Россия

Рассматривается проблема создания современных автоматизированных систем управления тематической информацией, разнородные по структуре источники которой размещены на территориально распределенной сетевой среде и представляют собой большие, корпоративного масштаба, или сверхбольшие (Интернет) хранилища данных. Предложены основанные на опытных данных новые подходы к управлению информацией включая механизмы, математические модели, алгоритмы и аппаратно-программные средства их реализации.

**Ключевые слова:** пространства данных, полнотекстовый поиск, графовые модели, полуструктурированные данные, виртуальная интеграция, качество услуг, онтология, разграничение доступа.

The article contains the views of author on creation of contemporary automated systems for managing thematic information, sources of which, heterogeneous in structure, are located on large distributed network environment and are large, corporate-sized and extra-large, Internet data storages. Ideas, presented by author, are based on new approaches to controlling information, including mechanisms, models, algorithms and hardware and software implementations, and on the experience of a group of researches, that he is representing.

**Key words:** data spaces, full-text search, graph models, half-structured data, virtual integration, quality of service, ontology, access control.

**Введение.** На современном этапе развития информационно-вычислительных и коммуникационных технологий, сетевой инфраструктуры проблема эффективного управления большими объемами доступных пользователям электронных данных является одной из наиболее актуальных. Здесь и далее под большими объемами понимаются такие объемы данных, которые по объективным причинам сложно хранить и использовать в сильноструктурированном виде (например, под управлением реляционных СУБД). Решение данной проблемы во многом определяет текущее состояние, развитие мировой цивилизации и каждого государства не только в ближайшем будущем, но и на многие годы вперед. Все основные сферы, характеризующие такое развитие в масштабах отдельных стран, а именно материальная, социальная, политическая и духовная (в том числе определяющая инновационный потенциал), в значительной степени зависят от решения задач в информационной сфере [1, 2]. По этой причине современное постиндустриальное общество принято называть информационным.

Рассматриваемая проблема является многофакторной, и ее успешное решение зависит от решения целого ряда сложных научно-технических, административно-организационных и практических задач, из которых базовыми, по мнению автора, являются следующие:

1. Управление мониторингом и поиском, систематизацией и анализом данных в больших (корпоративного масштаба) и сверхбольших (Интернет) коллекциях (хранилищах).
2. Управление качеством информационных услуг, предоставляемых в сетях передачи данных конечным пользователям.

3. Управление надежностью и безопасностью использования данных, сосредоточенных в больших и сверхбольших коллекциях. (Здесь безопасность понимается более широко, чем в традиционной безопасности информационных технологий, учитываются и социальное, и политическое значения (социальный, политический ущерб) от ее использования или деструктивных воздействий на нее.)

Решению данных задач посвящены исследования коллективов ведущих научных центров и университетов мира, о чем свидетельствуют цитируемые далее источники. Известны результаты в области разработки отдельных механизмов, формальных моделей и алгоритмов, формирующих основу математического обеспечения для решения каждой из перечисленных выше базовых задач. Существуют и находятся на разных стадиях жизненного цикла реализующие их аппаратно-программные средства. Предлагаются концептуальные положения подходов к объединению разрабатываемого математического и программного обеспечения в рамках новой абстракции (парадигмы) управления информацией на основе тематических пространств данных [3]. Однако в настоящее время не известны практически реализованные технические и технологические решения, доведенные как минимум до стадии экспериментального образца единой системы обработки, хранения, передачи и защиты тематической информации в слабоструктурированных пространствах данных, которые, во-первых, хотя бы на тестовых испытаниях демонстрировали возможность практического построения такой целевой системы и в какой-либо отдельной предметной области подтвердили бы заявляемые характеристики ее функциональных возможностей и качества их исполнения; во-вторых, показали бы инновационные перспективы подобной системы.

По мнению автора, исследования, проводимые в перечисленных выше направлениях, создание экспериментального образца такой системы, ее тестовые испытания и анализ их результатов являются приоритетной предметной областью на современном этапе развития методов и средств управления информацией.

**Пространство данных – предметная область и современная парадигма организации данных и управления ими.** Традиционные подходы к управлению данными с помощью их организации в виде баз данных (БД) с единой структурой (моделью) и системой управления (СУБД – система управления базой данных), эффективно применявшиеся с начала 1960-х гг., в настоящее время подвергаются серьезному пересмотру. Одной из главных причин такого положения дел является разработка в 70-80-е гг. XX в. технологий пакетных коммуникаций [4–6], а также создание метасети Интернет на основе стека протоколов TCP/IP [7] и бурное развитие "наложенной" на эту сеть информационной инфраструктуры Web на основе протокола HTML [8]. Таким образом, в метасети стали активно размещаться, пополняться и совместно существовать информационные активы не только в виде БД под управлением традиционных, как правило, реляционных СУБД, но и другие, так называемые слабо- или неструктурированные ресурсы, такие как Web, XML, аудио- и (или) видеоданные. В конце XX в. проблема управления разнотипными данными в электронном виде из источников в больших (корпоративного масштаба) и сверхбольших (Интернет) коллекциях (хранилищах) из разряда исследовательских, в качестве которой она рассматривалась в 1980-е гг., переросла в национально (транснационально) значимую и даже критически важную, во многом определяющую будущее человечества. Необходимость решения данной проблемы обусловлена прежде всего, тем что для эффективного функционирования любого достаточно крупного и сложноорганизованного хозяйствующего (а также действующего в социальной, духовной, политической сферах общественных отношений) субъекта (предприятия, корпорации, государственного ведомства) требуется перманентный анализ информации, относящейся к его направлению деятельности (тематический анализ) с целью принятия на этой основе своевременных и адекватных управленческих решений. Использование для этих целей традиционных, как правило, реляционных БД и СУБД под единым административным управлением в современных условиях в силу ряда объективных причин не представляется возможным. Однако следует отметить, что эволюция (модификация) таких СУБД, разработка и внедрение в практику новых механизмов, которые позволяли бы решать с их помощью некоторые из вновь возникающих задач, проводится.

В настоящее время в качестве ключевой программы действий, направленных на разрешение проблемы эффективного управления данными в большом числе разнородных, тематически связанных источников, рассматриваются проектирование, разработка и поэтапное внедрение в практику инструментальной платформы для поддержки пространств данных – DSSP (dataspace support platform) [3]. Пространством данных будем называть совокупность данных, размещаемых в различных источниках (корпоративных БД, на веб-сайтах и информационных порталах в

Интернете). Как правило, эти источники обладают различными схемами (моделями) организации данных, а иногда и не имеют таковых (структуры) вообще.

Представленная в настоящей работе (см. рисунок) классификация существующих и в той или иной степени апробированных решений при их размещении на плоской системе координат (по осям – высокий или низкий уровень семантической интеграции данных; близкая или далекая степень их административной соподчиненности или принадлежности) позволяет выявить следующие особенности.

1. Чем более "далеки" по уровню административной принадлежности (соподчиненности) источники данных, тем более развитыми в плане семантики (смысла) их описания должны быть средства запроса данных и манипулирования ими над группой источников, которые эти данные содержат.

2. Наиболее важными в рассматриваемом пространстве являются системы интеграции данных, предназначенные для поддержания механизмов объединения данных в источниках с разными схемами и моделями организации, запрашивания и манипулирования данными: неструктурированными (например, каталоги аудио- и видеоданных); текстовыми, представленными в html-формате; полуструктурированными (слабоструктурированными) в форматах OEM и XLM-моделей; строгоструктурированными под управление различных СУБД. Специфика подходов в этом направлении состоит в том, что при разработке механизмов интеграции данных в таких системах, как правило, необходим учет их семантики.

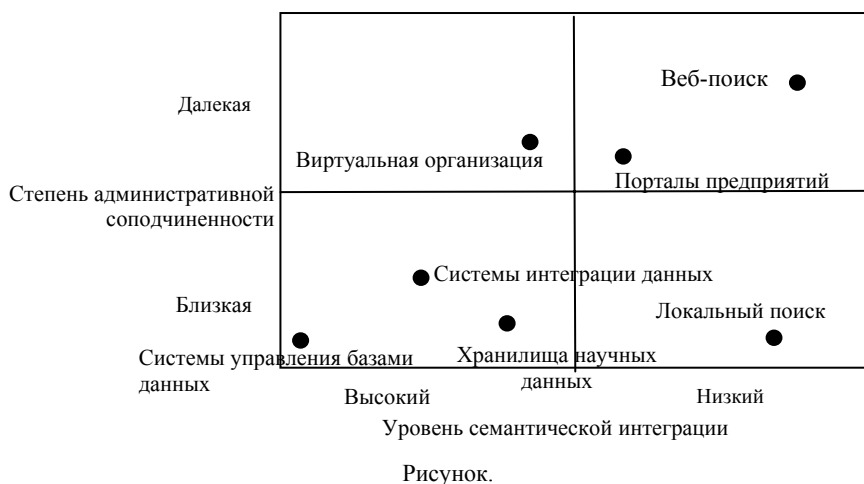
3. В целом подход к управлению данными на основе пространств данных не тождествен созданию систем интеграции. В настоящее время в дополнение к системам интеграции специалисты (эксперты) предлагают создавать базовые механизмы и модели, средства и системы, которые поддерживают представительный набор основных функций, необходимых для управления подобными данными. При этом в меньшей степени, по крайней мере на первом этапе, рассматриваемый подход призван обеспечивать качество и надежность выполнения этих функций.

Таким образом, DSSP должна представлять собой инструментальное средство (комплекс программ) для реализации более тесной интеграции источников данных и поддерживать механизмы, реализующие следующие функции:

- мониторинг изменения источников тематической информации в подконтрольном пространстве (корпоративном, региональном, в определенном, например национальном, сегменте Интернет) и их анализ с целью отбора экземпляров данных (объектов, документов);
- кластеризация экземпляров данных в семантически близкие коллекции (понятия);
- присваивание классам объектов и отдельным объектам соответствующих атрибутов и их индексирование;
- установление (описание) отношений между классами объектов и объектами внутри классов (например, в виде правил вывода – продукций, позволяющих судить о свойствах объектов, которые могут интересовать потенциальных пользователей).

В качестве источников, представляющих интерес для потенциального пользователя, должны рассматриваться текстовые данные в электронном виде, представленные в веб-среде Интернет, в ее национальных сегментах, имея в виду языковые возможности; полуструктурированные (слабоструктурированные) данные на основе OEM, XML и аналогичных моделей; строгоструктурированные данные на основе реляционной, объектно-ориентированной, других моделей и соответствующих СУБД; неструктурированные аудио- видеоданные.

С учетом отмеченных свойств инструментального комплекса для управления разноструктурированной тематической информацией, требований к его структуре модель пространства данных тематического содержания в его современном понимании представляет собой некоторое (первое) приближение онтологии предметной области,



которой эта тематика соответствует. Как отмечалось ранее, в концептуальном плане один из целесообразных подходов к управлению большими объемами тематических данных состоит в том, чтобы одновременно создавать средства и системы интеграции перечисленных выше источников информации. Суть данного подхода заключается в том, что различные типы таких источников (уровни интеграции) требуют разных затрат на внедрение средств управления содержащимися в них данными. Соответственно по мере изменения потребностей пользователи могут применять средства работы с данными источников как одного типа, так и другого. При таком понимании нижним уровнем интеграции в рамках модели пространств данных является уровень управления на основе механизмов контекстного (полнотекстового) поиска. Наличие на данном уровне общего поискового сервера, который индексирует все доступные информационные ресурсы, может рассматриваться как первый шаг к решению задачи интеграции данных, поскольку такой сервер является единым интерфейсом доступа ко многим ресурсам. Следует отметить, что затраты на внедрение такого решения минимальны, а во многих приложениях данный уровень интеграции является достаточным. Если возможностей управления информацией на основе механизмов контекстного поиска недостаточно, например вследствие ограничений языка запросов на использование логической структуры документов, то далее можно рассматривать подходы к управлению данными на основе графовых или древовидных моделей их представления [9].

Представление данных в виде графов с помеченными ребрами позволяет выразить структурные зависимости между объектами, не требуя при этом значительного изменения существующих схем хранения данных. Использование древовидных моделей приводит к более эффективным алгоритмам вычисления запросов, однако это может потребовать существенной модификации данных. Таким образом, реализация механизмов этого уровня интеграции связана с большими затратами на построение отображения данных из имеющейся модели в графовую. Вместе с тем рассматриваемый подход позволяет использовать возможности более богатых языков запросов.

На последнем уровне интеграции целесообразно использовать механизмы построения виртуальной распределенной базы с логически единой схемой данных. Несомненным преимуществом этого уровня является высокая релевантность результатов вычисления запроса. Например, выполнение SQL-запроса в реляционной базе данных, как правило, приводит к получению предсказуемого результата. Однако для построения такой системы необходимо решить задачу сопоставления схем данных в разных базах, которая не может быть решена в автоматическом режиме.

С учетом изложенного рассмотрим более подробно вопросы, которые возникают на каждом из перечисленных выше уровней интеграции данных, а также уже существующие и перспективные подходы к их разрешению.

**Управление на основе полнотекстового поиска в пространствах данных.** Полнотекстовый поиск, или поиск по ключевым словам, в силу объективных причин стал одним из первых подходов, которые позволяют в какой-то мере удовлетворять потребности, возникающие на новом с позиций современных требований и ставшем в последние десять лет научно значимым поле управления данными в электронном виде. В настоящее время в метасети Интернет с различной степенью эффективности эксплуатируется достаточно большое число подобных систем. Появление такого подхода обусловлено следующими объективными предпосылками:

- простота и доступность html-стандарта представления данных и манипулирования ими с помощью веб-браузеров, которые в значительной степени основаны на инженерно-эвристических подходах;
- высокие темпы внедрения данного подхода во все сферы жизнедеятельности общества;
- отсутствие, по крайней мере на первых этапах развития технологий полнотекстового поиска, необходимости применения строгих математических моделей, связанных с ними сложных алгоритмов и реализующего их программного обеспечения.

Однако, несмотря на активную работу в этом направлении крупнейших IT-компаний мира, высокую популярность таких зарубежных и российских систем, как Google и Yahoo, Yandex и Rambler, основными проблемами на пути их эффективного использования на практике остаются:

- низкая ревалентность поиска как следствие недостатков (должного уровня интеллекта) существующих алгоритмов кластеризации и ранжирования данных;
- относительно невысокие скорости поиска и анализа данных как результат присутствия еще достаточно "большой" доли неавтоматизированного "ручного" труда в едином цикле управления данными, что особенно проявляется при решении нетрадиционных в тематическом плане задач.

Первым серьезным препятствием на пути разрешения поставленных выше вопросов является большой объем тематической информации, которая распределена по различным источникам, сосредоточенным как в сетях корпоративного и ведомственного масштаба, так и в Интернете. Перманентный мониторинг (просмотр, обход) этих источников и выделение из них близких к определенной тематике данных для вторичной обработки и анализа представляет собой отдельную и технологически сложную задачу. Определенная работа в этом направлении проводится указанными выше мировыми компаниями – лидерами на рынке полнотекстового поиска [2–4], в том числе российскими коммерческими фирмами и исследовательскими центрами [10–12].

Второй причиной трудностей, которые приходится преодолевать на пути повышения релевантности поиска, является присутствие в языках:

- омонимов, когда одни и те же слова обозначают разные сущности (понятия, явления), например "ядро", "источник", "пакет", "политика";
- синонимов, разными словами описывающих одну и ту же сущность, например "позавчера администратор изменил настройки безопасности", "два дня назад ответственный за безопасность поменял параметры моделей разграничения доступа";
- слов с совпадающими словоформами, например "мой" – притяжательное местоимение и форма единственного числа повелительного наклонения глагола "мыть";
- слов, активно используемых в запросах с разными значениями сущностей, которые они обозначают в разных предметных областях, например "операция", "поток", "сеть".

Одним из способов преодоления перечисленных трудностей, который активно развивается в системах поиска по ключевым словам, является последовательное уточнение запроса и введение в него большего количества ключевых слов. Однако даже очень точно сформулированный запрос с большим количеством ключевых слов не позволит вычислить (определить) в источниках релевантные потребностям пользователя данные, которые описаны синонимами. В силу этого перспективным является указание, например, всех слов, которые могут быть синонимами слов исходных, адекватно характеризующих предметную область. Таким образом, появляется необходимость в описании семантики предметной области, в использовании элементов онтологии. В последние один-два года мировыми лидерами в области IT-технологий предлагаются некоторые, хотя и проприетарные (без открытых исходных кодов) программные продукты в указанных направлениях. Отдельные механизмы, подходы и программные средства в области математического и программного обеспечения решения указанных задач имеет коллектив исследователей Московского государственного университета им. М. В. Ломоносова [11,13].

Следует еще раз отметить, что интеллектуальные и ресурсные (материально-технические и человеческие) затраты на создание относительно полного, функционально замкнутого решения задачи управления данными на основе их поиска по ключевым словам не столь велики и для многих приложений в пространствах данных являются единственно оправданными.

**Управление данными на основе графовых моделей.** Следующим и более ресурсозатратным по сравнению с рассматриваемым выше на отмеченных уровнях иерархии источником тематической информации является уровень управления на основе графовых и древовидных моделей [14]. Как отмечалось ранее, данный подход может использоваться, если возможностей управления с помощью механизмов контекстного поиска оказывается недостаточно. Применяемая в этом подходе формальная модель данных представляет данные в виде ориентированного графа с помеченными ребрами. Она разрабатывалась для унификации формата передачи данных между разнородными приложениями [15]. Графическое представление данных обладает большой выразительной силой, поскольку в таком виде может быть представлена практически любая структура данных при условии правильного выбора множества меток. Однако практика показывает, что в этом случае поиск данных имеет высокую алгоритмическую сложность. Большинство языков запросов, предложенных для этой модели данных, приводит к необходимости решения NP-полных задач.

Следует отметить, что сложность вычисления запросов для графовых моделей данных и широкое распространение языка разметки XML во многом способствовали смещению исследований в последнее десятилетие в сторону древовидных моделей данных, использование которых приводит к более эффективным алгоритмам вычисления запросов, однако, как правило, требует существенного изменения схемы хранения данных. Вместе с тем результа-

ты, полученные за время многолетних исследований графовых моделей данных, в том числе коллективом, который представляет автор [10,16–19], позволяют надеяться, что такие модели будут востребованы на практике, если их дополнить элементами онтологий и метаданных. Основания для такого вывода следующие. Во-первых, онтологии отражают взаимосвязи между объектами и понятиями реального мира, что естественным образом приводит к возникновению ориентированных графов с помеченными ребрами. Во-вторых, графовая структура возникает при поиске в гипертекстовых документах. После идентификации в документах понятий, представленных в онтологии, между ними возникают дополнительные связи, обусловленные гипертекстовой структурой.

Модель полуструктурированных данных, как и модель реляционных данных, включает три составляющие: определение структуры данных, определение манипуляционной части и определение ограничений целостности. Первая составляющая описывает допустимую структуру объектов данных. Манипуляционная составляющая модели определяет систему операций над объектами. В число таких операций входят поиск, создание и модификация объектов. Ограничения целостности, представляющие собой совокупность логических утверждений, которым должны удовлетворять данные, делятся на определение схемы данных и определение дополнительных ограничений на них. Сложные структуры данных могут быть представлены в виде ориентированного графа с помеченными ребрами. Вершины графа соответствуют элементам структуры данных, а ребра представляют отношения между ними. Такое представление данных является достаточно универсальным.

Языки запросов к базам полуструктурированных данных проектируются с учетом их главной особенности – отсутствия единой строго определенной схемы. Семантически одинаковые фрагменты базы данных могут иметь различную структуру данных. Язык запросов должен содержать специальные механизмы, обеспечивающие получение "осмысленного" результата даже при несоответствии структуры данных в различных фрагментах базы данных. Так как отношения между объектами в рассматриваемой графовой модели данных выражаются путями между вершинами базы, то задача поиска сводится к нахождению вершин графа базы данных, связанных определенными путями. Среди методов описания критериев поиска можно выделить два существенно различающихся подхода, а именно функциональный поиск и поиск по шаблону. Функциональный подход выбран в системе UnQL [20], в которой запрос представляется в виде совокупности взаимно-рекурсивных функций специального вида (структурной рекурсии).

Языки запросов к полуструктурированным данным, основанные на поиске по шаблону (см., например, работу [21]), в качестве базового механизма используют регулярные путевые выражения, которые, по сути, являются регулярными языками. При этом пара вершин базы данных удовлетворяет регулярному путевому выражению, если между этими вершинами найдется по крайней мере один ориентированный путь, метки которого образуют слово из этого языка. Таким образом, регулярное путевое выражение, позволяет сформулировать условие, которому должны удовлетворять объекты, соответствующие вершинам базы данных.

В рассматриваемой модели данные представляются ориентированными графами, и схема данных должна определять некоторые множества ориентированных графов. Такие задачи, как проверка соответствия документа заданной схеме, определение вложенности схем или усечение запроса, должны быть эффективно разрешимы. Одним из подходов к описанию схем полуструктурированных данных являются графовые схемы [22]. Графовой схемой называется ориентированный граф с корневыми вершинами, ребра которого помечены унарными предикатами над некоторым алфавитом.

Рассмотрим основные задачи, возникающие при вычислении конъюнктивных регулярных путевых запросов. Поскольку база данных и запрос представляют собой ориентированные графы, а вычисление запроса требует нахождения отображения вершин запроса в множество вершин базы, задача вычисления запроса аналогична поиску подграфа в графе. В свою очередь, возможность эффективного вычисления запросов связана с решением ряда задач, которые традиционны по постановке для реляционных баз, однако имеют специфику для слабоструктурированных данных. Одной из таких задач является построение плана вычисления запроса. Основными стратегиями вычисления конъюнктивных регулярных путевых запросов CRPQ (conjunctive regular path query) являются исчерпывающий поиск и метод слияния результатов вычисления элементарных запросов. Первая стратегия, по сути, является алгоритмом поиска подграфа. Вторая стратегия предполагает вычисление элементарных запросов и последующее слияние результатов. Как показано в работах [17, 23], эффективность применения данных стратегий

существенно зависит от структуры вычисляемого запроса. Другой задачей является уменьшение (усечение) пространства поиска. Механизмы, модели и инструментальные средства, реализующие такой подход, представлены, например, в работах [23–26]. Процесс усечения пространства поиска на основании анализа набора схем документов основан на следующем факте: если известно множество схем, которым соответствуют все документы из базы данных, то, проверяя запрос на какой-либо одной из них, можно отсеять множество документов, на которых поиск по запросу заведомо не даст положительного результата.

Третьим направлением сокращения времени вычисления запроса является использование материализованных представлений. Материализованные представления, т. е. предварительно вычисленные результаты некоторых запросов, широко используются в современных базах данных для уменьшения времени вычисления запросов. Если система получает запрос, результат вычисления которого может быть получен на основании значений материализованных представлений, то часть вычислений, необходимых для получения результата, уже выполнена на этапе построения представлений. Такой подход приводит к сокращению общего времени вычисления запроса. Результаты, закладывающие математические основы решения перечисленных задач применительно к слабоструктурированным базам данных, получены в работах как зарубежных [27–31], так и российских [19, 29, 32–35] авторов.

Как отмечалось ранее, одним из перспективных подходов к более адекватному отражению семантических свойств данных, которые представляются графовыми моделями, является использование онтологий. Распространенным методом формального представления онтологий являются различные дескриптивные логики. Под онтологией в данном случае понимается пара  $K = \langle T, A \rangle$ , где  $T$  содержит универсальные утверждения, а  $A$  – утверждения о свойствах индивидуальных объектов. Интерпретация  $I$  является моделью онтологии  $K = \langle T, A \rangle$ , если она удовлетворяет всем утверждениям из  $A$  и аксиомам из  $T$  онтологии  $K$ . При этом интерпретация  $I$  онтологии пространства тематических данных может быть естественным образом представлена в виде ориентированного графа с помеченными ребрами.

Следовательно, для поиска экземпляров данных как объектов такой предметной области могут использоваться описанные ранее языки запросов. Однако заметим, что онтология может допускать бесконечное число моделей, и вычисление запроса к ней следует рассматривать вместе с задачей построения модели. Описанные методы вычисления запросов необходимо модифицировать для поиска гарантированных ответов, т. е. ответов, которые справедливы для любой интерпретации данной онтологии. Следует также отметить, что практическая реализация данного подхода связана с необходимостью решения ряда очень сложных задач включая построение строгих и содержательно адекватных математических моделей онтологий, эффективных алгоритмов и технологий, их программного обеспечения. Однако исследования в этом направлении представляются очень перспективными, в том числе в прикладном плане.

**Управление данными на основе систем виртуальной интеграции.** Еще одним традиционным типом (классом) источников информации, на основе интеграции которых может быть сформирован следующий уровень управления тематическими данными, являются строго структурированные базы данных (реляционные, объектно-ориентированные и др.) под управлением соответствующих СУБД. В этом случае решение задачи сводится к построению виртуальной распределенной базы данных с логически единой схемой их описания. Как отмечалось выше, преимуществом такой БД является релевантность результатов вычисления запросов, однако при реализации механизмов сопоставления схем локальных данных возникают объективные трудности.

Первые подходы к интеграции разнородных баз данных, обусловленные в большей степени исследовательскими целями, чем реальными потребностями в управлении данными, предлагались еще в 1980-х гг. В последнее десятилетие в связи с практической необходимостью в интеграции данных, ростом популярности новых механизмов и моделей полуструктурированных данных, развитием языка XML и стандартов работы с данными на его основе, в том числе языка запросов X Query, появились прототипы реальных систем виртуальной интеграции [36, 37].

Основным сектором практического использования систем виртуальной интеграции стали корпоративные системы автоматизированного управления. Задача на этом поле заключается в необходимости интеграции разнородных (в указанном выше смысле) баз данных, как правило, относящихся к одной предметной области. В качестве средства интеграции таких БД выступает набор механизмов, поддерживающих преобразование начальных запро-

сов к интегрированной (глобальной) БД, выработку плана его выполнения, конвертацию подзапросов к локальным БД, получение и обработку результатов, вычисление отдельных подзапросов и плана выполнения запроса в целом, выдачу итогового результата. В качестве локальных БД для систем виртуальной интеграции рассматриваются реляционные, XML/ xQuery базы данных. Вычисление результатов отдельных подзапросов к локальным БД обеспечивается за счет механизмов оберток (wrappers), а выполнение плана запроса реализуется с помощью совокупности механизмов, называемых посредниками – медиаторами (mediators). В рамках сервисно-ориентированной архитектуры виртуальной БД эти механизмы выполняются (оформляются) в виде сервисов (как правило, веб-сервисов).

Обертка как сервис, поддерживающий доступ к локальной БД, может представлять собой как логически простой адаптер, так и сложный в алгоритмическом плане интерфейс взаимодействия с ней, отражающий семантику данных в базе, учитывающий элементы онтологии предметной области. Еще более сложные в логико-алгоритмическом плане задачи приходится решать на этапе формирования плана выполнения запроса. Такой план реализуется на основе отображения схем локальных БД на схему единой извне виртуальной базы данных. На этом этапе с помощью механизмов оберток обеспечивается отображение схем локальных БД в глобальную схему с общим языком запросов к ней. Степень автоматизации процесса (процедуры) реализации такого отображения, его корректности, представительности набора локальных БД и скорости выполнения определяют интеллектуальный уровень виртуальной базы данных.

Следует отметить, что при разработке математического и программного обеспечения сервисов, формирующих планы выполнения запросов, используются все перечисленные атрибуты и методы исследования, характеризующие пространство тематических данных как отдельную предметную область с описанием ее онтологии. Основными элементами такого описания являются:

- объекты (экземпляры) предметной области в виде конечного набора входных и выходных параметров, которые используются для формирования запросов и результатов их выполнения;
- конечный набор шаблонов, кластеризующих запросы и результаты их вычисления;
- совокупность операций, формализующих отношения между шаблонами запросов и результатами их вычисления.

Подходы к решению перечисленных выше задач изложены в работе иностранных авторов [42]. Результаты исследований российских ученых в данном направлении представлены, например, в [36, 39], а также в работах [40, 41].

**Управление качеством информационных услуг, предоставляемых в пространствах данных.** Одной из важнейших проблем, которую необходимо решить для эффективного управления тематической информацией в пространствах данных, является создание системы управления качеством информационных услуг, предоставляемых в них конечным пользователям и приложениям. Создание такой системы опирается на возможности протоколов передачи данных в сетях, поддерживающих пространство данных, а также на производные от них механизмы манипулирования (обработки) данными. Системообразующим стеком таких протоколов, лежащих в основе метасети Интернет, корпоративных и ведомственных сетей, в настоящее время является TCP/IP. В силу этого в качестве одной из первоочередных мер по решению рассматриваемой проблемы выступает разработка протокольной базы для управления качеством обслуживания в IP-сетях, позволяющего обеспечивать "справедливое" (в соответствии с принятым в сети регламентом) разделение полосы пропускания между пользователями и приложениями.

В настоящее время традиционным для IP-сетей является принцип доставки, согласно которому все пакеты данных находятся в равных условиях. Это означает, что один пользователь или одно приложение может загрузить информационный канал, а негативные последствия (рост задержек, увеличение вероятности потерь пакетов) ощущают все пользователи и все приложения в равной степени. Следует отметить, что задача "справедливого" разделения полосы пропускания актуальна и в случае, когда выделяются специальные приложения, для которых осуществляется приоритетное обслуживание. Причина в том, что оставшуюся при этом пропускную способность каналов необходимо каким-либо разумным образом разделить между остальными потребителями. Для решения поставленной задачи необходимо разработать специальные дисциплины очередей для использования на промежуточных узлах; специальные алгоритмы управления потоками для магистральных соединений; протоколы для рас-



пространения информации между маршрутизаторами о желаемых (принятых) правилах классификации и приоритизации пакетов, предназначенных для конкретных потребителей.

Решение, включающее перечисленные выше механизмы, позволит интерактивным приложениям, наиболее чувствительным к качеству связи, получать приоритетное обслуживание даже без специальных настроек конечных или промежуточных устройств в сети. В случае специальных пожеланий пользователя оно позволит учесть их, если оператор связи располагает необходимыми для этого ресурсами. Разработанные протоколы, а также их реализации должны удовлетворять следующим требованиям.

1. Масштабируемость – протоколы должны быть пригодны к применению, в том числе на магистральных маршрутизаторах, не потребляя при этом слишком больших вычислительных ресурсов.

2. Децентрализованность – должна присутствовать возможность управления качеством обслуживания "из конца в конец" через сети различных операторов связи; при этом учитываются как потребности конечных пользователей, так и интересы, ресурсные возможности операторов.

3. Безопасность – никакой пользователь не должен иметь возможность глобально нарушить регламентированные режимы работы протоколов либо обеспечить себе несправедливо большую по сравнению со среднестатистической долю ресурсов.

4. Интероперабельность с имеющимся сетевым оборудованием – разработанные решения должны предоставлять возможность поэтапной замены оборудования, используемого оператором связи, с улучшением качества обслуживания на каждом этапе.

5. Автоматизация – протоколы должны обеспечивать повышение качества обслуживания даже при минимальных начальных настройках узлов сети, без наблюдения и коррекции настроек со стороны администратора.

Для разработки протоколов как с заявленными выше, так и с другими характеристиками, большое значение имеет возможность проведения имитационного моделирования. В настоящее время на рынке информационных технологий для проведения такого рода исследований представлен ряд программных средств, в том числе с открытым исходным кодом, например симулятор сетевых структур ns-2. Поэтому возникает необходимость выбора одного из их числа, его доработки и модификации, построения с помощью такого симулятора системы имитационного моделирования процессов маршрутизации трафика и передачи данных в сетях рассматриваемого класса. Доработки должны включать модули для поддержки разработанных алгоритмов и протоколов, а также средства трансляции высокоуровневых моделей сетей во входной язык симулятора.

В настоящее время известны идеи и подходы к разработке некоторых необходимых для решения поставленной выше задачи механизмов и протоколов, а именно очередей и алгоритмов управления потоками. Результаты их реализации представлены, например, в работах [42–46]. Однако комплексных моделей, позволяющих судить о количественных показателях их эффективности в сетях с реальными характеристиками, пока не разработано. Неизвестно также о каких-либо реализациях протоколов и алгоритмов управления потоками, которые были бы пригодны для натурных испытаний на экспериментальных полигонах.

На первом этапе научных и прикладных исследований в указанном направлении могут рассматриваться следующие.

1. Спецификация протоколов распределенного управления потоками, предназначенная для применения на магистральных сетях операторов связи.

2. Создание прототипа реализации разработанных протоколов и алгоритмов в виде программных модулей и модификаций ядра ОС Linux и программной системы для поддержки динамической маршрутизации Quagga. Прототип должен быть пригоден для натурных испытаний на полигоне, моделирующем магистральную сеть оператора.

3. Разработка математической модели передачи трафика в магистральной сети. Модель должна теоретически показывать преимущества разработанных алгоритмов и протоколов перед традиционными решениями, а также включать метрики качества обслуживания, актуальные в сетях нового поколения.

4. Создание комплекса программ для имитационного моделирования процессов маршрутизации и передачи данных в сетях, поддерживающего разработанные протоколы и алгоритмы наряду с традиционными протоколами семейства TCP/IP.

Перечисленные выше механизмы и математические модели, алгоритмы и программные средства призваны расширить существующие представления об информационном пространстве как предметной области и внести новые элементы в ее более строгое, формальное описание.

**Управление надежностью и безопасностью в пространствах тематических данных.** Рассматривая в качестве базовых описанные ранее уровни интеграции данных в тематических пространствах, необходимо учитывать, что по мере накопления сведений о таких данных (например, в виде описания их классов (типов), аннотаций, материализованных запросов), собственно элементах данных и отношениях между ними, увеличиваются их объем и ценность. Как следствие появляются дополнительные риски неблагоприятных событий, повышается уровень потерь, обусловленных, например, угрозами реализации деструктивных воздействий на данные (как умышленных, так и совершаемых не по злому умыслу). Возникает необходимость в разработке механизмов и построении формальных моделей, в создании реализующих их программных средств, которые могли бы с высокой степенью доверия гарантировать надежную работу пространства данных, защиту информационных активов и поддерживающих их средств инфокоммуникаций от деструктивных воздействий.

Важнейшей компонентой системы управления данными в тематическом пространстве, обеспечивающей высокий уровень надежности обслуживания пользователей, является подсистема мониторинга ее состояния на предмет оценки работоспособности. Такие подсистемы достаточно хорошо представлены на мировом рынке информационных технологий [47–52]. Большинство таких систем поставляются пользователям без открытых исходных кодов, что создает дополнительные трудности при их модификации под постоянно (по объективным причинам) меняющиеся потребности столь же быстро эволюционирующих технологий управления данными в тематическом пространстве. Однако имеются такие средства и с открытыми кодами, которые, как правило, разрабатываются научными коллективами в рамках открытых проектов. Отметим, например, работы зарубежных [51, 52] и российских [53, 54] исследователей.

Основным вектором развития средств мониторинга состояния больших и сложноорганизованных систем, подобных рассматриваемым в настоящей работе, является их интеллектуализация и автоматизация. Исследования и практические работы в этом направлении ориентированы на создание механизмов, поддерживающих изменения в конфигурации системы; обнаружение "сбоев" в работе составляющих ее элементов; оперативное реагирование на такие "сбои" в автономном режиме, с высоким уровнем его автоматизации. Решения в этом направлении предлагают учитывать семантику тематического пространства данных как предметной области, а именно используемого коммуникационного оборудования и средств вычислительной техники, а также технологию анализа потенциально возможных событий и формирование мер оперативного реагирования. В силу этого важной представляется задача достаточно детальной формализации рассматриваемой предметной области, которая включала бы описание классов используемых объектов и связей между ними, для того чтобы на основе их анализа в автоматизированном режиме принимать адекватные решения.

Задача организации безопасного использования информационных ресурсов тематического пространства данных кроме традиционных мер, механизмов и сервисов, основанных на комплексном подходе к ее решению, имеет определенную специфику, обусловленную тем, что в ходе мониторинга информационных источников, анализа, кластеризации и других действий с тематическими данными происходит перманентная модификация пространства данных, появляются элементы с новыми характеристиками, в том числе категориями доступа к ним. Последнее обстоятельство должно приводить к изменениям в модели логического разграничения доступа к данным, в которой появляются новые объекты с соответствующими им метками доступа. Анализ новых элементов пространства данных на предмет характеризующих их категорий (меток) доступа и модификация на этой основе модели логического разграничения доступа представляет собой отдельную сложную задачу. Решение этой задачи в автоматизированном режиме должно опираться на анализ семантики вновь появляющихся элементов данных. Результаты исследований в этом направлении представлены, например, в работе зарубежных ученых [55], относящейся к "контентной фильтрации", и работе российских авторов [56]. В последние 2–3 года такие исследования проводятся с участием студентов и аспирантов на механико-математическом факультете и в Институте проблем информационной безопасности Московского государственного университета им. М. В. Ломоносова (см., например, [1, 2, 57, 58]).

Дополнительные механизмы защиты данных могут обеспечиваться путем использования указанных выше материализованных представлений (вычисленных ранее запросов). Пользователю могут, например, предоставляться права на доступ к данным только через представления. В этом случае он не имеет доступа к хранящимся в той же базе данным, которые согласно принятой политике безопасности не предназначены для него. Согласно положениям той же политики ответы на некоторые другие запросы могут быть конфиденциальными и для этой категории пользователей. Более того, описанные механизмы могут также потребовать обоснования того, что доступные пользователю представления не позволяют ему вычислять такие конфиденциальные запросы. Необходимо уметь проверять, принадлежит ли какой-либо запрещенный пользователю запрос множеству запросов, которые могут быть вычислены на основании разрешенных ему представлений. В более общей форме задачу можно ставить как проверку пустоты пересечения множества запрещенных запросов с множеством запросов, которые вычисляются на основании набора разрешенных представлений. Необходимо отметить, что данная задача является актуальной для автоматизированных информационных систем, поддерживающих объекты с повышенными требованиями к их безопасности [2]. Для решения этой задачи могут использоваться результаты исследований, полученные в работах [16, 59, 60].

**Заключение.** Изложенные в настоящей работе материалы не описывают набора проектных решений, направленных на создание полномасштабной системы управления тематической информацией, источники которой размещены на большом по масштабам и гетерогенном (разнородном) по составу поддерживающей их аппаратно-программной среде пространстве данных. Представленные идеи в большей степени определяют отправные посылы, современные требования к подобной системе и соображения по поводу перечня и содержания научных исследований и опытно-конструкторских работ, критический анализ результатов выполнения которых позволил бы перейти к этапу ее проектирования.

Необходимо еще раз отметить, что представленные выше подходы включая механизмы, модели, алгоритмы и их аппаратно-программные реализации являются новыми. В настоящее время они находятся в центре внимания научно-технических коллективов ведущих университетов и научных центров мира. Перспективность этих научных и прикладных исследований подтверждается многочисленными публикациями, часть которых представлена в списке цитируемых источников. О возможности практической реализации подходов, предлагаемых в данной работе, свидетельствуют результаты исследований, аппаратно-программные реализации и результаты тестовых испытаний, выполненных участниками научного коллектива, который представляет автор. В этой связи автор выражает благодарность А. С. Козицыну, С. А. Афонину, А. С. Шундееву, В. Б. Савкину, принимавшим активное участие в подготовке настоящей публикации.

### Список литературы

1. АНДРЕЕВ О. О., ВАСЕНИН В. А. и др. Критически важные объекты и кибертерроризм. Ч. 1. Системный подход к организации противодействия / Под ред. В. А. Васенина. М.: МЦНМО, 2008.
2. АНДРЕЕВ О. О., ВАСЕНИН В. А. и др. Критически важные объекты и кибертерроризм. Ч. 2. Аспекты программной реализации / Под ред. В. А. Васенина. М.: МЦНМО, 2008.
3. FRANCLIN M., HALEVY A., MAIER D. From databases to dafepases: a new abstraction for information monedement // SIGMOD Record. 2005. V. 34, N 4. [Electron resource]. [http://www.citforum.ru/databe/articles/from\\_ab\\_to\\_ds/](http://www.citforum.ru/databe/articles/from_ab_to_ds/).
4. КЛЕЙНРОК Л. Вычислительные системы с очередями. М.: Мир, 1979.
5. МИЗИН И. А. Сети коммутации пакетов / И. А. Мизин, В. А. Богатырев, А. П. Кулешов. М.: Радио и связь, 1986.
6. БЛЕК Ю. Сети ЭВМ: протоколы, стандарты, интерфейсы. М.: Мир, 1990.
7. ВАСЕНИН В. А. Российские академические сети Интернет (состояние проблемы решения) / Под ред. В. А. Садовниченко. М.: РЭФИА, 1997.
8. САДОВНИЧИЙ В. А., ВАСЕНИН В. А. и др. Российский Интернет в цифрах и фактах / В. А. Садовнический, В. А. Васенин, А. А. Мокроусов. М.: Изд-во Моск. гос. ун-та, 1999.
9. ВАСЕНИН В. А. К созданию концепции интегрированной информационной системы распределенных информационных ресурсов Московского государственного университета им. М. В. Ломоносова / В. А. Васенин, С. А. Афонин, А. А. Коршунов. М.: Изд-во Московск. гос. ун-та, 2001.

10. ВАСЕНИН В.А., АФОНИН С.А., КОЗИЦЫН А.С., ШУНДЕЕВ А.С. Поиск в сверхбольших хранилищах данных и высокопроизводительные системы с массовым параллелизмом // Тр. Междунар. конф. "Программные системы: теория и приложения". Переславль-Залесский, 13–14 мая 2004. М.: Физматлит, 2004. Т. 1. С. 211–228.
11. ВАСЕНИН В. А., АФОНИН С. А., КОЗИЦЫН А. С. Автоматизированная система тематического анализа информации // Информационные технологии. 2009. № 4, (Приложение).
12. МАЖУКИН А. В. Поисковые системы сети Internet. М.: Изд-во Моск. гуманитарного ун-та, 2007.
13. АФОНИН С. А., КОЗИЦЫН А. С., ТИТОВ А. С. Автоматизированная система для тематического анализа информации // Критически важные объекты и кибертерроризм. Ч. 2. Аспекты программной реализации средств противодействия / Под ред. В. А. Васенина. М.: МЦНМО, 2008. С. 497–577.
14. VASENIN V. A., AFONIN S. A. To the problem of building an integrated system of university distributed information resources // Proc. of the Finnish data processing week conf., Petrozavodsk, 2001. Petrozavodsk State University, 2001. P. 152–177.
15. PAPA-KONSTANTINOU Y., GARCIA-MOLINA H., WIDOM J. Object exchange across heterogeneous information sources // Proc. of the Intern. conf. on data engng. IEEE Computer Society, Washington (USA), 1995. P. 251–260.
16. AFONIN S., KHAZOVA E. Membership and finiteness problems for rational sets of regular languages // Intern. J. Foundations Computer Sci., 2006. N 17(3). P. 493–506.
17. АФОНИН С. А. Алгоритмы эффективного вычисления конъюнктивных регулярных путевых запросов // Вычисл. технологии. 2007. № 12(2). С. 24–33.
18. Afonin S. The view selection problem for regular path queries // Proc. of the 9th Intern. symp. on theoretical informatics LATIN 2008, Buzios, 2008. Lecture Notes in Computer Science. Berlin, 2008. V. 4957. P. 121–132.
19. ХАЗОВА Е. Е. Алгоритмы построения однословной перезаписи регулярных путевых запросов. Дис. ... канд. физ.-мат. наук. М., 2009.
20. BUNEMAN P., DAVIDSON S., HILLEBRAND G., SUCIU D. A query language and optimization techniques for unstructured data // ACM SIGMOD Records, 25(2), ACM, New York, 1996. P. 505–516.
21. АВТЕВОУЛ С., ВИАУ В. Regular path queries with constraints // Proc. of the Intern. symph. on principles of database systems (PODS 97), 1997. P. 122–133.
22. FERNANDEZ M. F., SUCIU D. Optimizing regular path expressions using graph schemas // Proc. of the 14th Intern. conf. on data engng. Orlando (USA), IEEE Computer Society Press, Washington, DC, USA, 1998. P. 14–23.
23. АФОНИН А. С. Алгоритмы эффективного вычисления конъюнктивных регулярных путевых запросов в базах полуструктурированных данных. Дис. ... канд. физ.-мат наук. М., 2007.
24. ГОРЕЛОВ С. С., ВАСЕНИН В. А. Усечение пространства поиска в полуструктурированных базах данных при помощи иерархии схем документов // Программирование. 2005. № 6. С. 41–55.
25. ГОРЕЛОВ С. С. Оптимальные иерархии схем для поиска по конъюнктивным регулярным путевым запросам в полуструктурированных базах данных // Программирование. 2006. № 4. С. 38–56.
26. ГОРЕЛОВ С. С. Эффективные модели поиска в базах полуструктурированных данных на основе иерархии схем документов. Дис. ... канд. физ.-мат наук. М., 2009.
27. HIRST G. Ontology and the Lexicon: Handbook on ontologies in information systems. Springer, Series on Handbooks in Information Systems, S. Staab, R. Studer (eds.). Springer, 2003. P. 209–230.
28. CALVANESE D., EITER T., ORTIZ M. Regular path queries in expressive description logics with nominals // Proc. of the 21st Intern. joint conf. on artificial intelligence (IJCAI 2009), Pasadena (USA), Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 2009. P. 714–720.
29. CALVANESE D., GIUSEPPE D.G., LENZERINI M., ROSATI R. View-based query answering over description logic ontologies // Proc. of the 11th Intern. conf. on the principles of knowledge representation and reasoning (KR 2008), Sydney, Australia, AAAI Press, 2008. P. 242–251.
30. КУЗНЕЦОВ С. Д., ПЛЕШАЧКОВ П. О. Управление транзакциями в РСУБД с поддержкой XML // Программирование. 2006 № 5. С. 3–20.
31. COOX S., SIMANOVSKY A. Regular expressions in XML Schema Evolution // Proc. of the ISTA'2003. Kharkiv (Ukraine), Lecture Notes in Informatics, GI-Edition, 2003. P. 1–7.
32. КУЗНЕЦОВ С. Д., АЗНАУРЯН Н. А., НОВАК Л. Г., ГРИНЕВ М. Н. SLS: нумерующая схема для больших XML-документов // Программирование. 2006. № 1. С. 12–25.
33. KUZNETSOV S., FOMICHEV A., GRINEV M.. Sedna: A Native XML DBMS // Proc. of the 32nd Conf. on current trends in theory and practice of computer science. Lecture notes in computer science. Merin (Czech Republic), 2006. Lecture notes in computer science. V. 3831. Berlin, 2006. P. 272–281.

34. LUKICHEV M., BARASHEV D. XML Query algebra for cost-based optimization // Proc. of the SYRCoDIS, Moscow, 2007. CEUR, Aachen, 2007. P. 62–69.
35. АФОНИН С. А., ГОРЕЛОВ С. С., ХАЗОВА Е. Е. Слабоструктурированные данные и поиск на основе онтологий // Материалы Всерос. конф. с междунар. участием "Знания-онтологии-теории" (ЗОНТ-09), 22–24 окт. 2009 г. Новосибирск, Ин-т математики им. С. Л. Соболева СО РАН, 2009. С. 119–123.
36. АНТИПИН К. В., ФОМИЧЕВ А. В., ГРИНЕВ М. Н. и др. Оперативная интеграция данных на основе XML: системная архитектура BizQuery // Тр. Ин-та системного программирования. 2004. № 5. С. 157–174.
37. FERNANDEZ M., KADYSKA Y., SUCIU D. SilkRoute: A framework for publishing relational data in XML // ACM Transactions on Database Systems. 2002. N 27(4). P. 438–493.
38. CHAWATHE S., GARCIA-MOLINA H., HAMMER J., ET AL. The TSIMMIS project: integration of heterogeneous information sources // Proc. of the 10th Meeting of the information processing society of Japan, Tokyo (Japan), Oct., 1994. Information Processing Society of Japan, Tokyo, 1994. P. 7–18.
39. КАЛИНИЧЕНКО Л. А. Методы и средства интеграции неоднородных баз данных / Под ред. Л. Н. Королева. М.: Наука. Глав. ред. физ.-мат. лит., 1983.
40. ШУНДЕЕВ А. С., ПЕРШИН И. С. Интеграция данных: от баз данных к информационным ресурсам // Материалы Всерос. конф. с междунар. участием "Знания – онтологии – теории" (ЗОНТ – 09), 20–22 окт. 2009. Новосибирск, Институт математики им. С. Л. Соболева СО РАН, 2009. С. 41–45.
41. МАКУНИН И. В. К вопросу о построении системы виртуальной интеграции данных для эффективного управления разнородными данными // Информационные технологии моделирования и управления. 2008. Вып. 1(44). С. 78–86.
42. JORDI ROS, WEI K. TSAI. An optimal distributed protocol for fast convergence to maxmin rate allocation // Department of Electrical and Computer Eng., Univ. of California, Irvine (USA). June, 2000. [Electron resource]. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.37.1444>.
43. HARRISON D., YONG XIA, SHIVKUMAR KALYANARAMAN, KISHORE RAMACHADRAN. An edge-based framework for flow control // ESCE and CS Departments, Rensselaer Polytechnic Institute, 2002. [Electron resource]. <http://www.ecse.rpi.edu/Homepages/shivkuma/research/papers/edge-based.pdf>.
44. SALIL S. KANHERE, HARISH SETHU. On the latency bound of pre-order deficit round robin // Proc. of the IEEE 27th Annual Conf. on local computer, networks. Nov., 2002, Tampa (USA). P. 508–517.
45. SALIL S. KANHERE, HARISH SETHU. Prioritized elastic round robin: an efficient and low-latency packet scheduler with improved fairness // Techn. Rep. DU-CS-03-03, Department of Computer Sci., Drexel Univ., Philadelphia (USA). 2003. PA 19104.
46. ВАСЕНИН В. А., СИМОНОВА Г. И. Математические модели управления трафиков в Интернет: новые подходы на основе TCP/AQM // Автоматика и телемеханика. 2005. № 8. С. 948–960.
47. DROGSETH D. EMA: network lifecycle management. A solution approach to managing networks. [Electron resource]. [http://www.harbrook.net/consulting/solutions/pdf/Downloads/EMA\\_HP-NLM\\_WP.pdf](http://www.harbrook.net/consulting/solutions/pdf/Downloads/EMA_HP-NLM_WP.pdf).
48. up.time IT systems management software. [Electron resource]. <http://www.uptimesoftware.com/overview.php>.
49. FREY J., KNAPP L., SZULEWSKI R. Network management. [Electron resource]. (<http://www.redbooks.ibm.com/redpapers/pdfs/redp3676.pdf>).
50. [Electron resource]. [http://en.wikipedia.org/wiki/Comparison\\_of\\_network\\_monitoring\\_systemsT](http://en.wikipedia.org/wiki/Comparison_of_network_monitoring_systemsT).
51. BARTH W. Nagios, 2nd edition system and network monitoring. No Starch Press, Oct. 2008. P. 720.
52. TARUS BALOG. Enterprise-Wide Network Management with OpenNMS. [Electron resource]. <http://www.oreillynet.com/pub/a/sysadmin/2005/09/08/opennms.html>.
53. ВАСЕНИН В. А., КОРНЕЕВ В. В., ЛАНДИНА М. Ю., РОГАНОВ В. А. Система функционального активного мониторинга FLAME. Программирование. 2003. № 3. С. 57–72.
54. АСТАПОВ И. С., ДЗЫБА М. С., КОРШУНОВ А. А. Система мониторинга работоспособности информационно-телекоммуникационной среды // Тез. Междунар. молодежной науч.-техн. конф. "Высокопроизводительные вычислительные системы" (ВПВС-2008), Таганрог (Россия), 1–5 сент. 2008 г. С. 5–7.
55. JOSEM GOMEZ IGNACIO, JOSÉ M. GÓMEZ, IGNACIO GIRÁLDEZ, MANUEL DE BUENAGA. Text categorization for internet content filtering // Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. 2003. № 20. P. 34–52.
56. ЗАБОРОВСКИЙ В. С., СИЛИНЕНКО А. В. Логико-динамические аспекты моделирования процессов контентной фильтрации прикладных протоколов // Материалы III Междунар. науч. конф. по проблемам безопасности и противодействия терроризму, Москва (Россия), 25–27 окт. 2007 г. М.: МЦНМО, 2008. С. 272–277.
57. ВАСЕНИН В. А., ШАПЧЕНКО К. А., АНДРЕЕВ О. О. Математические модели и механизмы логического разграничения доступа в операционной системе Linux: текущее состояние и перспективы развития // Материалы II Междунар. науч. конф. по про-

- блемам безопасности и противодействия терроризму // V Общерос. науч. конф. "Математика и безопасность информационных технологий" (МаБИТ-06), Москва (Россия), 25–26 окт. 2006 г. М.: МЦНМО, 2007. С. 159–171.
58. ШАПЧЕНКО К. А. К вопросу о средствах ОС Linux для управления доступом при использовании ролевых политик безопасности // Материалы Междунар. науч. конф. по проблемам безопасности и противодействия терроризму // IV Общерос. науч. конф. "Математика и безопасность информационных технологий". Москва, 2–3 ноября 2005 г. М.: МЦНМО, 2006. С. 257–281.
59. AFONIN S., KHAZOVA E. Membership and finiteness problems for rational sets of regular languages // Proc. of the Developments in language theory / Ed. by C. De Felice, A. Restivo. Lecture Notes in Computer Science. S. 1. V. 3572. 2005. P. 88–99.
60. ГРИНЕВА М. П. Методы поддержки активного поведения систем управления базами XML-данных. Дис. ... канд. физ.-мат. наук. М.: 2007.

*Васенин Валерий Александрович – д-р физ.-мат. наук, проф. Московского гос. ун-та;  
зав. лаб. Науч.-исслед. ин-та механики МГУ;  
зав. отделом Ин-та проблем информационной безопасности МГУ; e-mail: vassenin@msu.ru*

Дата поступления – 05.11.2009 г.