

ОРГАНИЗАЦИЯ ПАКЕТНОЙ СИНХРОНИЗАЦИИ РЕЛЯЦИОННЫХ БАЗ ДАННЫХ

Дмитрий Викторович Жучков, к.т.н., научный сотрудник

Тел.: 8 391 249 4834, e-mail: dimkas@icm.krasn.ru

ФГБУН Институт вычислительного моделирования

Сибирского отделения Российской академии наук

http://icm.krasn.ru

Представлен подход к решению задачи синхронизации структуры и информационного наполнения реляционных баз данных, используемых в территориально распределенной сети организаций. Решение основано на унификации и метаструктуризации объектов базы данных, а также использовании XML-пакетов, передаваемых через транспортный веб-узел.

Ключевые слова: базы данных, пакеты данных, таблицы, оглавления, синхронизация, транспортный узел, XML.

*Работа выполнена при поддержке гранта
РФФИ № 16-07-01001*

Введение и описание задачи

Задача синхронизации реляционных баз данных возникает при необходимости обеспечения идентичности структуры и информационного наполнения нескольких экземпляров базы данных, эксплуатируемых в организациях, которые включены в многосторонний информационный обмен. Периодическая синхронизация баз данных обеспечивает возможность совместной работы пользователей из различных организаций с массивами данных и электронными документами.

Примером задачи, для которой описываемый подход был впервые применен, является сбор и согласование данных в процессе планирования и осуществления закупок для муниципальных нужд города Красноярска.



Д.В. Жучков

Муниципальные учреждения и подразделения администрации Красноярска образуют трехуровневую иерархическую структуру, внутри которой идет интенсивный информационный обмен в рамках задачи муниципальных закупок. Организационная структура включает более 350 муниципальных учреждений, 27 главных распорядителей бюджетных средств (департаментов и управлений) и 1 уполномоченный орган, выполняющий роль управляющего центра системы. Задачей муниципальных учреждений является формирование заявок на закупки, главные распорядители бюджетных средств согласуют заявки и формируют документацию по закупкам, а уполномоченный орган согласует документацию и проводит все закупочные процедуры [1].

На момент разработки и внедрения нового подхода к автоматизированному обмену данными сформировался ряд базовых условий и ограничений, которым разрабатываемое программное обеспечение должно удовлетворять:

1. Участники системы должны сохранить собственные независимые базы данных, работа с которыми осуществляется по локальной вычислительной сети;
2. Отправка и получение данных от других участников должны выполняться с использованием сети «Интернет» непосредственно из системы, без применения промежуточных файловых форматов и средств электронной почты;
3. Средства информационного обмена должны обладать гибкими настройками для оперативного изменения схемы хранения и передачи данных;
4. Необходимо обеспечить автоматическое обновление всех составляющих си-

стемы, включая структуру пользовательских данных, программное обеспечение, а также централизованные справочники и классификаторы.

В результате выполненных работ в описанной организационной структуре обеспечен многосторонний информационный обмен, при котором каждый из участников имеет собственный экземпляр реляционной базы данных, а передача данных между участниками выполняется путем обмена информационными сообщениями с использованием промежуточного транспортного веб-узла.

В результате каждого частного случая обмена информационными пакетами происходит частичная синхронизация принимающего экземпляра базы данных с отправителем. В результате обмена вышестоящие организации имеют у себя полные копии баз данных подведомственных организаций, а уполномоченный орган, отвечающий за функционирование всей системы в целом, обладает полным набором данных, полученных от всех участников системы. При этом вышестоящие организации имеют возможность как самостоятельно корректировать полученные данные, а затем передавать их вниз по иерархии в исправленном виде, так и возвращать присланную информацию назад для корректировки в исходное учреждение.

В перечень передаваемых и согласуемых данных входят табличные сведения о закупках для муниципальных нужд, а также электронные версии документации по закупкам. Реализованный процесс электронного согласования документации позволяет существенно сократить трудозатраты на внесение изменений в текст документации, а также ускоряет получение итоговых версий документов, подписанных электронными подписями.

Для реализации информационного обмена проведены подготовительные работы в части унификации и метаструктуризации объектов баз данных системы, а также разработано специализированное программное обеспечение, работающее на стороне участников и на центральном сервере системы. Далее в работе описаны основные составляющие решения представленной задачи.

Унификация и метаструктуризация объектов базы данных

Для повышения эффективности управления базой данных системы и реализации средств информационного обмена, обладающих возможностями гибкой настройки, выполнена предварительная унификация объектов базы данных. За основу унификации взята авторская информационная модель унифицированного справочника [2].

В результате проведенной унификации все таблицы, составляющие базу данных, обладают схожей структурой, задаваемой следующими общими правилами построения таблиц.

1. Каждая таблица имеет целочисленное автоинкрементное ключевое поле, которое используется для установки ссылок на записи этой таблицы в других таблицах в пределах одного экземпляра базы данных. Целочисленный ключ используется для упорядочения записей в порядке их создания и хорошо воспринимается пользователями при поиске необходимой информации.

2. Таблицы, участвующие в многостороннем обмене, имеют альтернативный ключ, в виде глобального уникального идентификатора (globally unique identifier, GUID). Использование идентификаторов типа GUID позволяет обеспечить точность и уникальность ссылок на конкретную запись в рамках всей распределенной системы.

3. Каждая таблица содержит поля периода действия записи, в которых содержится информация о дате создания записи, дате и основании введения записи в действие, дате и основании прекращения действия записи, а также дате и основании последнего изменения записи.

4. Каждая таблица имеет отдельный журнал изменений, в котором фиксируются операции создания, изменения и удаления записей с полной копией состояния записи на момент события.

5. Все идентификаторы таблиц и полей в базе данных подчинены единой дисциплине имен. Поля таблиц, имеющие одинаковую структуру и назначение в различных таблицах, имеют одинаковые фиксированные идентификаторы. Идентификаторы целочисленных ключевых полей и стандартных полей «Наименование» формируются из идентификатора таблицы. Идентификаторы полей, являющихся ссылками на другие таблицы, включают в свой состав идентификаторы родительских ключевых полей.

Для обеспечения гибкости настроек системы хранения и передачи данных выполнена метаструктуризация объектов базы данных, т.е. в состав базы данных введены метаданные – служебные таблицы, в которых хранится дополнительная информация об объектах базы данных, их структуре и взаимосвязях между таблицами. Метаданные расширяют стандартный набор сведений о таблицах реляционной базы данных и формируют метаструктуру базы – адаптированное для нужд пользователей системы визуальное представление перечня таблиц, составляющих базу данных, и связей между ними.

В состав метаданных входят следующие сведения:

- структурное деление групп таблиц по разделам предметной области;
- сведения о таблицах системы, включая их локальный и глобальный идентификаторы, наименование, тип таблицы, информация о владельце, основание создания и период действия таблицы;
- структура полей таблицы – наименование, идентификатор, тип данных, обязательность заполнения, ссылка на другую таблицу, дополнительные характеристики;
- сведения о связях между таблицами, включая отношения зависимости и подчиненности;
- информация об особенностях передачи и обработки сведений из отдельных таблиц.

Визуальное представление основных метаданных в виде дерева объектов системы обеспечивает наглядный инструментарий для работы с объектами базы (рис. 1).

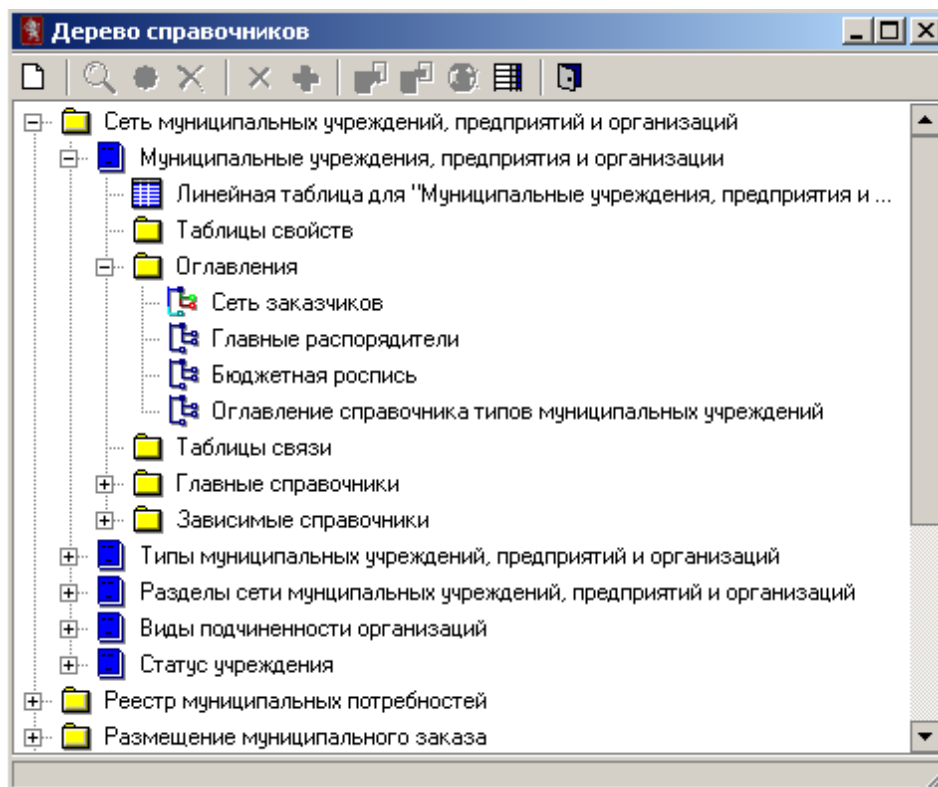


Рисунок 1. Древоподобное представление списка объектов базы данных

Хранение сведений об объектах базы данных внутри нее самой позволяет оперативно вмешиваться в структуру базы. При этом передача сведений об изменениях

структуры объектов базы данных передаётся теми же средствами, что и пользовательская информация.

Средства генерации информационных пакетов

Для реализации многостороннего обмена данными в базу данных введен дополнительный раздел служебных справочников, который содержит информацию о субъектах и объектах обмена данными, а также его особенностях. Этот служебный раздел содержит следующую информацию:

- перечень таблиц, включенных в многосторонний обмен, – для каждой таблицы задаются: идентификатор таблицы, способ обмена данными, ответственная организация;
- перечень точек обмена данными – список учреждений, участвующих в информационном обмене – для каждого учреждения задаются: идентификатор, тип учреждения, роль учреждения, вышестоящая организация;
- пользователи системы – для каждого пользователя указываются: ФИО, организация, роль, права доступа, параметры авторизации на транспортном узле;
- собственно передаваемые данные – пакеты – для каждого пакета задаются: тип пакета, наименование, идентификаторы, отправитель, адресат, описание передаваемых данных.

Ключевой особенностью перечня таблиц, участвующих в многостороннем обмене, является дополнительная иерархическая структура, определяющая вложенность данных при формировании исходящего информационного пакета (рис. 2).

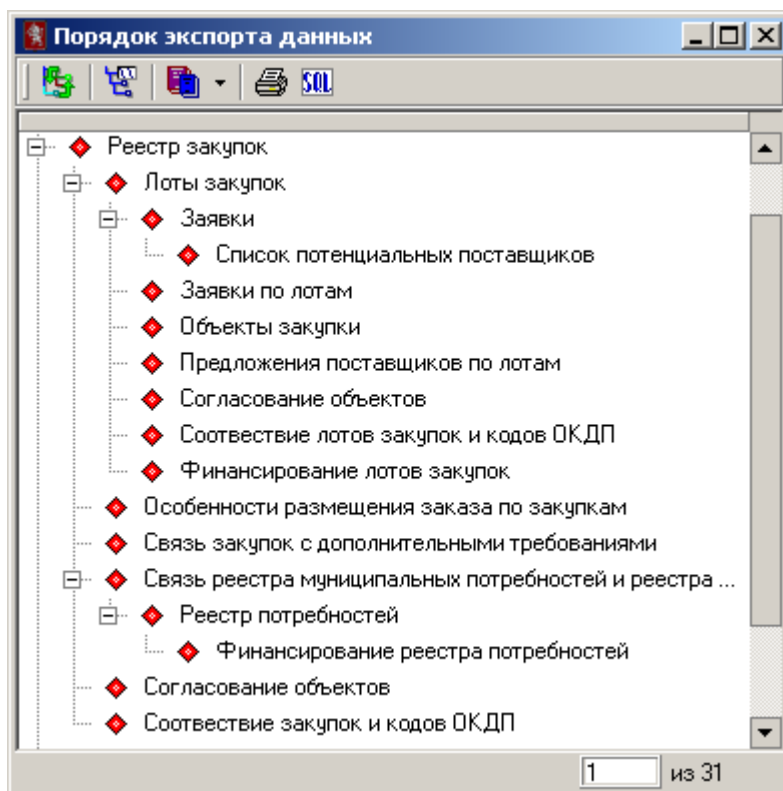


Рисунок 2. Структура сведений реестра закупок

Как видно из рис. 2 в состав сведений реестра закупок, находящегося на первом уровне дерева, включаются 16 дополнительных таблиц – лоты закупок, заявки по лота, сведения о согласовании объектов и другие. При формировании исходящих пакетов с данными программное обеспечение, отвечающее за формирование пакета, анализирует приведенную выше структуру и выполняет выборку данных в соответствии с установленными настройками. При этом визуализация вложенности данных наглядно показывает структуру информационного пакета и позволяет оперативно ее изменять.

На вход модуля генерации пакетов подаётся выборка ключевых значений таблицы первого уровня, а наборы вложенных данных формируются рекурсивно в соответствии с условиями связей между таблицами. Основная часть связей производится через ключевые поля, а для особых случаев предусмотрено хранение специальных правил и дополнительных условий выборки подчиненных данных. На выходе модуля генерации пакетов создаётся XML-структура утвержденного вида, содержащая все необходимые сведения.

Сформированная XML-структура упаковывается в пакет-контейнер, содержащий дополнительную заголовочную информацию, в которой указываются: отправитель, адресат, тип пакет, версия протокола, дата создания и т.д. При необходимости передаваемые в пакете данные могут быть сжаты, а также для них может быть сформирована верифицирующая электронная подпись. Готовые пакеты сохраняются в базе данных системы, а затем передаются в транспортный модуль, который передаёт их адресату.

Пакетный веб-транспорт

Роль транспортного модуля системы выполняет специализированный веб-сервис `transport`, взаимодействие с которым осуществляется по протоколу `http`. Веб-сервис обладает защищенной сервис-ориентированной архитектурой и обеспечивает передачу информационных пакетов между участниками системы [3, 4].

Основными функциями транспортного модуля являются адресная, безадресная и многоадресная передачи пакетов, а также регистрация новых организаций и пользователей. Регистрация в веб-сервисе новых организаций и пользователей обеспечивает разделение доступа к передаваемым данным. Пользователи системы авторизуются в сервисе с автоматической привязкой к соответствующей организации. Запросы пользователей к сервису выполняются с учётом установленных на транспортном сервере ограничений доступа к данным.

Адресная передача пакетов является средством доставки пользовательских данных, предназначенных для конкретного получателя. Адресный пакет доставляется только отдельному получателю. Безадресная рассылка пакетов применяется при необходимости широковещательной рассылки пакетов, например, пакетов с системными обновлениями и обновлениями централизованных справочников и классификаторов.

Многоадресная рассылка пакетов имеет две разновидности – групповая рассылка и рассылка по подписке. Групповая рассылка применяется при необходимости доставки пакета всем учреждениям, принадлежащим к отдельной группе, например, только организациям одной отрасли или одного района. Рассылка по подписке является самым гибким инструментом передачи данных. Адресатом пакета в этом случае является синтетический объект типа «поток». Пользователи транспортного узла могут присоединяться к потоку, и тогда им приходят все пакеты, адресованные в этот поток. Рассылка по подписке особенно эффективна в случае необходимости передачи определенной информации заранее неизвестной группе получателей.

Примером передачи пакетов по подписке является обмен данными по муниципальным закупкам. Получателями потока пакетов, связанного с одной закупкой, автоматически назначаются уполномоченный орган, главный распорядитель бюджетных средств и муниципальное учреждение, создавшее закупку. При участии в одной закупке нескольких учреждений они также добавляются в подписку. В результате каждый пакет, направляемый в поток, связанный с этой закупкой, получают все заинтересованные стороны.

Использование веб-сервиса в качестве транспортного модуля позволяет пользователям системы, находящимся в различных организациях, оперативно обмениваться данными без дополнительных затрат.

Особенности применения входящих пакетов

Этап получения и применения информационных пакетов является наиболее важным в процессе синхронизации баз данных, так как именно на этом этапе обеспечивается целостность и непротиворечивость данных в системе в целом.

В рамках представляемого подхода каждый из экземпляров синхронизируемой базы данных поддерживает локальную ссылочную целостность данных. Это значит, что при установке связей между таблицами в базе используются механизмы внешних ключей, и СУБД контролирует, что каждая из ссылок ведёт на соответствующую запись в родительской таблице. В ходе формирования информационного пакета все входящие в него данные также сохраняют свою ссылочную целостность, но на принимающей стороне ссылочная целостность передаваемых данных не может быть сохранена без дополнительной коррекции данных, т.к. принимающая сторона обладает собственными ключами.

Таким образом, возникает задача корректного переноса данных, сформированных в одном экземпляре системы, в другой экземпляр с сохранением корректных связей внутри данных. Эта задача относится только к тем разделам базы данных, которые формируются пользователями на местах и затем передаются в другие организации для согласования и исправления. Централизованно распространяемые справочники и классификаторы сохраняют ключевые значения во всех базах данных и не требуют дополнительных манипуляций.

Примером данных, задействованных в многостороннем обмене, является информация о закупках, включающая электронные версии документов. Впервые данные о закупке формируются в муниципальном учреждении, и там же в систему помещаются первые версии электронной документации. Затем информация направляется на согласование главному распорядителю бюджетных средств и/или в уполномоченный орган. В ходе согласования данных в них вносятся изменения, при этом правом внесения изменений обладает любой из участников обмена и, соответственно, новые данные могут появиться на любом этапе обработки. Задачей системы на этом этапе является корректная передача всем заинтересованным сторонам полного пакета информации, с возможностью последующей его корректировки.

В качестве решения этой задачи использован механизм автоматического согласования ключей на принимающей стороне, основанный на использовании глобальных идентификаторов. Все табличные данные и электронные документы, участвующие в многостороннем обмене, снабжаются глобальными уникальными идентификаторами типа GUID. При поступлении пакета в принимающую базу данных все данные размещаются в специальных буферных таблицах. Буферные таблицы повторяют структуру рабочих таблиц системы, но не имеют включенных механизмов ссылочной целостности. После загрузки данных выполняется специальная процедура согласования ключей, которая выполняет замену ключевых значений, поступивших из базы отправителя пакета, на локальные ключи по условию совпадения глобальных идентификаторов. Впервые поступившие данные при этом получают новые локальные ключевые значения. После выполнения процедуры согласования ключей импортированные данные переносятся в рабочие таблицы системы.

Важной особенностью этапа передачи данных между организациями является перенос прав редактирования данных в составе пакета. Отправитель данных самостоятельно принимает решение о том, следует ли передать права на редактирование данных и кому именно их передать. В отдельных случаях права на внесение изменений никому не передаются, и тогда пакет имеет чисто уведомительное значение. В других случаях, когда права необходимо передать, это в явном виде указывается в пакете путём заполнения полей «статус объекта» и «текущее местонахождение объекта». Сочетание этих полей позволяет системе принять решение о наличии у каждого пользователя тех или иных прав. При этом становится невозможной ситуация, когда одни и те же данные управляются несколькими пользователями одновременно.

Автоматизация пакетного обмена данными

Для отдельных задач, решаемых с помощью средств пакетной синхронизации, актуально создание средств автоматической генерации пакетов и их рассылки адресатам по заранее установленным правилам. К таким задачам относятся, например, периоди-

ческие рассылки обновлений справочников и классификаторов, а также дежурные уведомления пользователей о передаче закупок, находящихся на согласовании, между отделами уполномоченного органа. Аналогичная ситуация складывается с автоматическим получением и применением пакетов, т.к. большинство пакетов могут быть применены без контроля со стороны пользователя.

Для решения этих задач на центральном сервере системы внедрены автоматические средства обработки входящих пакетов, а также очереди заданий на отправку пакетов. Входящие пакеты, направленные в уполномоченный орган, автоматически загружаются с сервера и, в зависимости от типа пакета, либо применяются, либо откладываются для ручного применения. В свою очередь, обработка заданий на отправку пакетов позволяет периодически формировать исходящие пакеты. В параметрах каждого задания на отправку пакета фиксируются тип пакета и его основные параметры. Механизмы автоматической генерации пакетов, получив задание, формируют необходимый пакет, определяют его получателей и отправляют пакет на транспортный сервер.

Развитие средств автоматической обработки пакетов позволяет существенно сократить время ожидания пользователей при получении пакетов, поскольку делает этот этап незаметным для пользователя.

Заключение

Представленный в работе подход к организации синхронизации баз данных с использованием унифицированных объектов базы данных и средств оперативного изменения форматов передаваемых пакетов, позволил организовать гибкий механизм многостороннего обмена данными между экземплярами базы данных, используемыми в более чем 100 организациях.

Программный комплекс автоматизированной поддержки процессов планирования и размещения муниципального заказа, реализующий представленный подход к синхронизации баз данных, с 2013 года внедрен в подразделениях администрации города Красноярска и в подведомственных учреждениях. За 2015 год через систему прошли согласование 2640 муниципальных закупок, включающих 19380 электронных документов. При этом через транспортный узел системы было передано 20470 пакетов, общим объемом более 16.5 Гб.

В настоящее время продолжаются работы по развитию программного комплекса, направленные на оптимизацию средств синхронизации данных, а также модернизацию механизмов системы в соответствии с изменениями нормативно-правовой базы и пожеланиями пользователей.

Литература

1. Жучков Д.В., Исаева О.С., Лузан Н.Ф., Ноженкова Л.Ф., Щербенин В.Ф. Комплексная автоматизированная поддержка подготовки, размещения и контроля муниципальных заказов // Материалы X Всеросс. научно-практической конф. «Проблемы информатизации региона». В 2 т. Т. 1. – Красноярск: Сиб. федер. ун-т; Политехн. ин-т, 2007. С. 3-11.
2. Виноградов К.А., Жучков Д.В., Никитина М.И. Система ведения базы данных нормативно-справочной информации // Врач и информационные технологии. – М.: Менеджер здравоохранения, 2004, №.3.
3. Кононов Д.Д. Организация защищенных веб-сервисов для решения задач муниципального управления города Красноярска // Материалы XIV Всероссийской конференции «Проблемы информатизации региона». – Красноярск: ИВМ СО РАН. 2015. С. 110-115.
4. Кононов Д.Д. Обеспечение безопасности в системе пакетного веб-транспорта // Материалы XVIII Междунар. науч. конф. "Решетневские чтения". – Красноярск: СибГАУ, 2014. Ч. 2. С. 314-316.

Organization of packet synchronization in relational databases

Dmitry Viktorovich Zhuchkov, candidate of technical sciences, scientific researcher, Institute of Computational Modeling Siberian Branch of Russian Academy of Sciences

The article presents an approach to solving of the problem of structural and informational synchronization in relational databases used in territorial distributed network of organizations. The presented solution is based on the unification and metastructuring of database objects, as well as using information XML-packets sent through the transitional transport web-node.

Keywords: relational database, data packet, tables, directory, reference, synchronization, transport web-node, XML.

УДК 004.9

ПРОБЛЕМЫ ИЗУЧЕНИЯ ПРОЦЕССОВ САМООРГАНИЗАЦИИ В ВЕБ-ПРОСТРАНСТВЕ

Ольга Анатольевна Клименко, канд. физ.-мат. наук., старший научный сотрудник
Тел.: +7 383 330 0564, e-mail: klimenko@ict.nsc.ru
Институт вычислительных технологий СО РАН
<http://www.ict.nsc.ru>

Роман Юрьевич Федоров, канд. филос. наук., старший научный сотрудник
Тел.: +7 (3452) 688-787, e-mail: r_fedorov@mail.ru
ФГБУН Институт криосферы Земли СО РАН
<http://www.ikz.ru>

Статья посвящена проблемам развития междисциплинарных подходов к изучению принципов самоорганизации веб-пространства и закономерностей существующих внутри него коммуникативных связей. В качестве методологической основы исследования использован синтез биосоциальных трактовок коммуникативных процессов и их моделирования с помощью вебометрики и теории графов.

Ключевые слова: интернет-сообщества, веб-пространство, вебометрия, теория графов.



О.А. Клименко

На протяжении 2012-2014 гг. в Сибирском отделении РАН был реализован междисциплинарный интеграционный проект по теме "Исследование закономерностей и тенденций развития самоорганизующихся систем на примере веб-пространства и биологических сообществ". Данный проект объединил коллективы 12 научно-исследовательских институтов, осуществляющих исследования не только в области информационных технологий, но и в таких сферах науки, как биофизика, генетика, органическая химия, социология и др. Подобная междисциплинарная консолидация была не случайной. Во время разработки концепции проекта, представителями разных научных дисциплин неоднократно отмечалось, что развитие самоорганизующихся систем различной природы подчиняется общим закономерностям, которые могут быть описаны с помощью математических моделей. К примеру, для описания процессов, происходящих в информационно-телекоммуникационных системах, используются биологические термины «всемирная паутина», «вирусы», «черви» и т.д. Фактически это указывает на эмпирически-интуитивное понимание того факта, что многие закономерности развития информационных отношений в технике и в живой природе носят универсальный характер [1].



Р.Ю. Фёдоров