

ИТ-ИНФРАСТРУКТУРА НАУЧНЫХ ИССЛЕДОВАНИЙ: МЕТОДИЧЕСКИЙ ПОДХОД И РЕАЛИЗАЦИЯ*

Л. В. МАССЕЛЬ, Е. А. БОЛДЫРЕВ, Н. Н. МАКАГОНОВА,
А. Н. КОПАЙГОРОДСКИЙ, А. В. ЧЕРНОУСОВ
*Институт систем энергетики им. Л.А. Мелентьева СО РАН,
Иркутск, Россия*
e-mail: massel@isem.sei.irk.ru, e.boldyrev@gmail.com,
mak@isem.sei.irk.ru, digger@istu.edu

The article is concerned with an approach to realization of IT infrastructure proposed by the authors. The approach is based on the service oriented architecture, web-services, metadata repository and ontologies.

Введение

Предлагаемый авторами подход к созданию ИТ-инфраструктуры¹ научных исследований является обобщением многолетних работ в области построения информационных технологий для системных исследований энергетики и интеграции их в рамках ИТ-инфраструктуры этих исследований [1]. В настоящее время можно выделить общие тенденции в области программного и информационного обеспечения научных исследований — это переход к удаленному использованию вычислительных, информационных и других ресурсов на основе концепций GRID и метакомпьютинга (обычно в виде таких ресурсов выступают какие-либо уникальные программные, аппаратные или программно-аппаратные комплексы, базы данных и др.). В свою очередь, процесс организации научных исследований эволюционирует в сторону виртуальных территориально распределенных групп, работающих над одним проектом, что обуславливает необходимость обеспечения одинаковой доступности информационных и вычислительных ресурсов проекта для всех его участников. Революционный переход от существующего положения вещей к современным реалиям в области как программного обеспечения, так и организации научных исследований невозможен ввиду отставания технологической базы. В то же время очевидна актуальность соответствующих методических разработок и научных прототипов, с тем чтобы

*Работа выполняется при частичной финансовой поддержке Российского фонда фундаментальных исследований (грант № 07-07-00265), РГНФ (грант № 07-02-12112).

¹ИТ-инфраструктура — инфраструктура научных исследований, базирующаяся на современных информационных технологиях.

© Институт вычислительных технологий Сибирского отделения Российской академии наук, 2006.

создать предпосылки для перехода к работе научных сотрудников в новой, современной ИТ-инфраструктуре.

Актуальность создания ИТ-инфраструктуры научных исследований определяется следующими факторами:

- необходимостью интеграции современных информационных технологий, используемых для научных исследований, с целью повышения их эффективности и расширения сферы применения результатов исследований, в том числе с помощью сети Интернет;

- важностью проблемы развития информационных и телекоммуникационных технологий в России, о чем свидетельствует появление проектов “Электронная Россия” (в масштабах страны), “Информационно-телекоммуникационные ресурсы СО РАН” (Сибирское отделение РАН) и др.;

- мировыми тенденциями развития информационных технологий (использование сети Интернет не только как связующей и информационной среды, но и как среды программирования; создание вычислительных и информационных ресурсов в сети Интернет, распространение технологий web-служб и web-сервисов и др.).

1. Фрактальный подход к созданию ИТ-инфраструктуры

Под *ИТ-инфраструктурой* понимается *телекоммуникационная распределенная вычислительная инфраструктура*, а именно совокупность технических и программных средств, телекоммуникаций и информационной базы научных исследований; технологий их разработки и использования; стандартов, как внутренних, так и внешних, для создания информационных и программных продуктов в области научных исследований, обмена ими и их представления на информационный рынок [2]. В составе ИТ-инфраструктуры выделяются: распределенная вычислительная, информационная и телекоммуникационная инфраструктуры (последняя связана с техническими аспектами создания ИТ-инфраструктуры и в данной статье не рассматривается).

Концепция создания ИТ-инфраструктуры методологически обосновывается с помощью фрактальной стратифицированной модели (ФС-модели) информационного пространства [3]. ФС-модель позволяет описать информационное пространство, в которое отображается вся имеющаяся информация по данной предметной области, в виде совокупности непересекающихся слоев, объединяющих однотипные информационные объекты (обладающие одинаковым набором свойств, или характеристик). Каждый слой, в свою очередь, может быть расслоен.

Использование ФС-модели позволяет представить информационную технологию как совокупность информационных слоев и их отображений, тогда инструментальные средства информационной технологии включают средства описания информационных слоев и средства поддержки отображений из любого слоя в каждый.

Графически фрактально-стратифицированная модель ИТ-инфраструктуры (рис. 1) может быть представлена в виде совокупности вложенных сферических оболочек (слоев), определяемой тройкой $\{S, F, G\}$, где S — множество слоев; F — множество отображений; G — множество инвариантов. ИТ-инфраструктура S , согласно фрактальной методологии, расслаивается (стратифицируется) на интеграционную информационную инфраструктуру (S_I), распределенную вычислительную инфраструктуру (S_C) и телекоммуникационную инфраструктуру (S_T). В свою очередь, S_I расслаивается на слои данных и метаданных (S_{ID}, S_{IM}), S_C — на слои программ и метаописаний программ (S_{CP}, S_{CM}). Вводятся отоб-

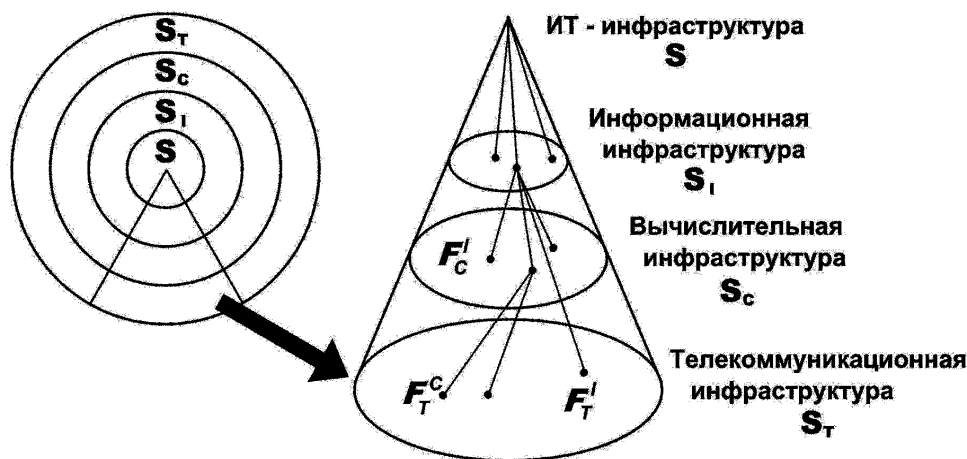


Рис. 1. Фрактальная стратифицированная модель ИТ-инфраструктуры.

ражения слоев: $F_C^I : S_I \rightarrow S_C$, $F_T^I : S_I \rightarrow S_T$, $F_T^C : S_C \rightarrow S_T$. Инвариантами (G) являются цели научных исследований, которые детализируются для каждого слоя. Для описания метаслоев предлагается использовать информационные модели, модели данных и онтологии, для реализации средств поддержки отображений — web-сервисы [2].

2. Методические основы и базовые технологии создания ИТ-инфраструктуры

Первоочередное внимание авторами уделяется разработке методических основ создания ИТ-инфраструктуры научных исследований в Интранет-Интернет [2, 4, 5]. Такая инфраструктура, с одной стороны, является интеграционной информационной и вычислительной средой для проведения научных исследований. С другой стороны, возникают предпосылки для поэтапного перехода к созданию распределенных баз данных и программных комплексов, распределению и распараллеливанию вычислений, созданию web-ориентированных программных комплексов (вычислительных серверов), оказанию информационных услуг на основе наукоемких информационных и программных продуктов (создание web-служб и web-сервисов).

В основе разработки ИТ-инфраструктуры лежат концепции GRID (Global Resource Information Distribution) и SOA (Service-Oriented Architecture). Суть идеи GRID состоит в том, что пользователю предлагаются не просто вычислительные мощности, распределенные в сети, а возможность использования программных компонентов и/или информационных баз для решения определенного класса задач, с которыми одновременно может работать большое количество специалистов [6, 7].

Формальное определение SOA сформулировано специалистами корпорации IBM: “SOA — это прикладная архитектура, в которой все функции определены как независимые сервисы с вызываемыми интерфейсами. Обращение к этим сервисам в определенной последовательности позволяет реализовать тот или иной бизнес-процесс” [8]. Обе концепции (GRID и SOA) направлены на интеграцию и повышение доступности имеющихся вычислительных, информационных и аппаратных ресурсов. GRID предполагает объединение имеющихся ресурсов в единую систему, а SOA — нет. Основная идея SOA — предоставление

функциональности приложений в виде сервисов и обеспечение доступа к ним независимо от используемых операционных платформ и языков программирования.

Рассмотрим, как согласуются основные положения разработки ИТ-инфраструктуры с концепцией SOA. Одно из ключевых понятий в SOA — *интероперабельность*, т. е. способность двух или более информационных систем или их компонентов к взаимодействию. Выделяют *семантическую интероперабельность* (свойство информационных систем, обеспечивающее взаимную употребимость полученной информации на основе общего понимания ее значения), и *техническую интероперабельность* (совместимость систем на техническом уровне, включая протоколы передачи данных и форматы представления данных). Техническая интероперабельность обеспечивается применением интеграционных технологий². Подходы к обеспечению семантической интероперабельности пока не стандартизованы. Наиболее популярные в настоящее время подходы предполагают, что, оперируя стандартными онтологиями (т. е. описаниями предметных областей), информационные системы смогут самостоятельно идентифицировать и связываться со всеми необходимыми им для выполнения требуемых задач компонентами, в ходе работы адаптируясь к техническим интерфейсам и преобразуя форматы данных.

Сервис-ориентированная архитектура предлагает разработчикам другой подход к многократному использованию кода: вместо традиционного объектно-ориентированного наследования предполагается композиция, т. е. создание более сложных сервисов из сервисов низкого уровня. При этом преодолевается основное ограничение наследования — сервисы могут быть распределены в сети и даже принадлежать разным организациям. Разработаны основополагающие стандарты и спецификации для создания web-сервисов [9].

Выделяют четыре этапа (уровня) перехода к SOA:

- реализация отдельных web-сервисов (технологии web-сервисов используются для разработки новых приложений или преобразования старых);
- сервисно-ориентированная интеграция бизнес-функций, в том числе интеграция на уровне пользовательского интерфейса, информационная интеграция, поддержка различных способов коммуникации низкого уровня между приложениями, интеграция процессов, интеграция унаследованных систем;
- трансформация ИТ-инфраструктуры в масштабе предприятия;
- изменения в деятельности предприятия [10].

В проекте, выполняемом авторами по созданию ИТ-инфраструктуры системных исследований в энергетике, получены результаты, которые могут быть сопоставлены с двумя первыми этапами. Ведутся работы в рамках третьего этапа [1, 2, 4, 5].

Проект выполняется на основе методического подхода, частично описанного в [1], основными составляющими которого являются:

1) *методические принципы создания распределенной вычислительной инфраструктуры*, включающие рекомендации по построению расширяемых программных комплексов (ПК), их объектному моделированию с использованием языков моделирования и CASE-средств, применению базовых технологий, создающих предпосылки для перехода к работе с ПК через Интернет; реализации ПК в трехуровневой архитектуре “клиент-сервер”, включающей сервер приложений и сервер баз данных;

2) *методика адаптации унаследованного программного обеспечения* и создания на его основе для исследований энергетике программных комплексов нового поколения, отличительные черты которых — свойство расширяемости программного обеспечения в его

²Таких как DCOM (Distributed Component Object Model), CORBA (Common Object Request Broker Architecture), RMI (Remote Method Invocation).

современной трактовке, многослойная архитектура, предусматривающая использование серверов приложений и серверов баз данных, ориентация на работу в вычислительных сетях (как локальных, так и глобальных);

3) *методические принципы создания информационно-инфраструктуры*, включающие определение основных требований к инфраструктуре, построение метаописаний данных и онтологий для описания систем энергетики, стандарты представления данных на основе XML;

4) *методические принципы и технология экспертной поддержки удаленных пользователей специализированных вычислительных серверов.*

3. Распределенная вычислительная инфраструктура

Методический подход к созданию распределенной вычислительной инфраструктуры был апробирован при разработке распределенного программного комплекса для исследований проблемы энергетической безопасности ИНТЭК³. При реализации этого программного комплекса созданы предпосылки для организации web-доступа и создания вычислительного сервера в сети Интернет. Этот же подход использовался при создании вычислительного сервера для решения задач оптимального управления на базе ПК OPTCON⁴.

На основе определяемого SOA взаимодействия компонентов web-сервисов предлагается архитектура распределенного программного комплекса, представленная на рис. 2.

Основные компоненты архитектуры:

— *вычислительное ядро*. Реализует основную функциональность программного комплекса;

— *посредник*. Отвечает за преобразование запросов от пользователей в вызовы методов (функций), реализованных в вычислительном ядре, и преобразование и передачу результатов их выполнения компонентам, отвечающим за взаимодействие с пользователями;

— *web-сервис*. Предоставляет доступ к функциональности комплекса (всей или частично, определяется интерфейсом сервиса) по протоколу SOAP;

— *клиенты*. Имеют доступ к функциональности комплекса по другим протоколам. Например, клиент, напрямую взаимодействующий с вычислительным ядром, может работать локально, а клиент, взаимодействующий с компонентом “посредник”, — по сети. Данные способы взаимодействия ориентированы на непосредственное использование комплекса людьми, а web-сервис — это способ взаимодействия с программным комплексом, ориентированный скорее на программное взаимодействие.

Предлагаемый подход позволяет предоставлять разным пользователям различные функции программного комплекса. Например, для пользователей, осуществляющих доступ через web-сервис (внешние пользователи), возможно предоставление ограниченного круга услуг, в то время как клиенты, использующие программный комплекс по другим протоколам, могут получать доступ к полной функциональности.

Следует отметить, что данный подход имеет смысл использовать там, где необходимы организация взаимодействия между разнородными программными комплексами или

³Реализация ПК ИНТЭК выполнена Е.А. Болдыревым, реализация сервера INTEC — А.П. Демьянчиком и Е.А. Болдыревым под руководством Л.В. Массель.

⁴Реализация ПК OPTCON выполнена А.Ю. Горновым, реализация вычислительного сервера — Д.В. Подкаменным и А.Ю. Горновым в рамках гранта Российского фонда фундаментальных исследований № 02-07-90343 (2002–2003 гг.) под руководством Л.В. Массель.

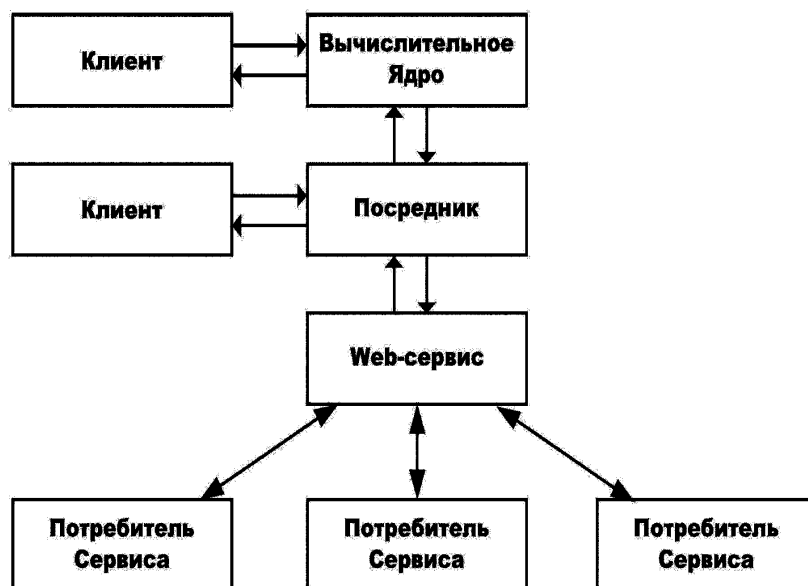


Рис. 2. Предлагаемая архитектура распределенного ПК в соответствии с концепцией SOA.

обеспечение доступа к ним для внешних пользователей. В других случаях может быть целесообразно применение более “легких” (с точки зрения необходимости развертывания дополнительной инфраструктуры) технологий.

4. Информационная инфраструктура

Особое место в ИТ-инфраструктуре занимает *интеграционная информационная инфраструктура*. Интеграция предполагает не физическое, а “виртуальное” объединение данных путем создания их единообразного описания (построения метаданных) с последующим созданием репозитория (хранилища метаданных) и принятие единого стандарта для организации обмена данными. Не менее важная задача при создании информационной инфраструктуры — описание имеющихся интеллектуальных ресурсов. Сейчас констатируется, что сеть Интернет должна обладать качествами “мышления”, из которых наиболее важны работа с предметными областями как с онтологиями и возможность переопределения системы понятий онтологии с целью упрощения процесса решения задачи. Онтологии сейчас — одно из наиболее интенсивно развиваемых направлений. Под онтологиями понимают базы знаний специального вида, которые могут “читаться” и пониматься, отчуждаться от разработчика и/или физически разделяться их пользователями. Это формально представленные знания на базе концептуализации (описания множества объектов и понятий, знаний о них и связей между ними) [11].

Информационная инфраструктура научных исследований поддерживает выполнение следующих функций [5]:

— создание и организация хранения метаданных (описаний) ресурсов института, которые можно подразделить на: информационные (базы данных, хранилища данных, наборы данных, онтологии); вычислительные (программы и программные комплексы); трудовые (научные сотрудники института, разработчики баз данных и программных продуктов); научно-исследовательские (НИР, проекты, гранты и др.); сервисные (web-сервисы, имеющиеся в институте); электронные (публикации, отчеты, архивы);

- поиск необходимого ресурса по метаданным;
- преобразование метаданных из внутреннего стандарта в общепринятые стандарты (например, DublinCore, RDF и др.);
- создание и организация хранения онтологий как в целом по предметной области, так и по ее разделам, а также по научным исследованиям, ведущимся в институте;
- организационная, информационная и вычислительная поддержка научных исследований.

Для хранения метаданных предлагается использовать репозитарий — систематизированное хранилище информации, которое включает базу метаданных, информационные ресурсы и системную информацию. База метаданных содержит описания баз данных, программных комплексов, проводимых исследований; описания онтологий; данные о сотрудниках или сторонних пользователях, участвующих в исследованиях; ссылки на конкретные информационные ресурсы.

Основные компоненты информационной инфраструктуры (рис. 3):

- клиентское приложение — так называемый “легкий клиент”, приложение, ориентированное на web-доступ к серверу и работающее в среде браузера (Интернет-проводник) с установленной поддержкой языка программирования Java;
- информационные сервисы — небольшие программы, работающие на сервере и предоставляющие такие услуги, как доступ к метаданным клиентским приложениям и другим сервисам, авторизация пользователей на сервере метаданных, преобразование метаданных в форматы открытых стандартов (XML, RDF, DublinCore и т. п.), отображение запрашиваемой информации из метабазы данных и управление структурой метаданных;
- репозитарий.

Для описания информационных ресурсов первоначально была использована инструментальная система Protege-2000, которая позволяет разработать онтологию предметной области, т. е. определить основные термины (концепты) предметной области, с помощью фреймов описать эти концепты через совокупность слотов. Эта система имеет возможность сохранить разработанную онтологию в одном из стандартных форматов (XML, RDF и RDFS, OWL и др.), используемых для обмена метаданными или базами знаний. В по-



Рис. 3. Основные компоненты информационной инфраструктуры.

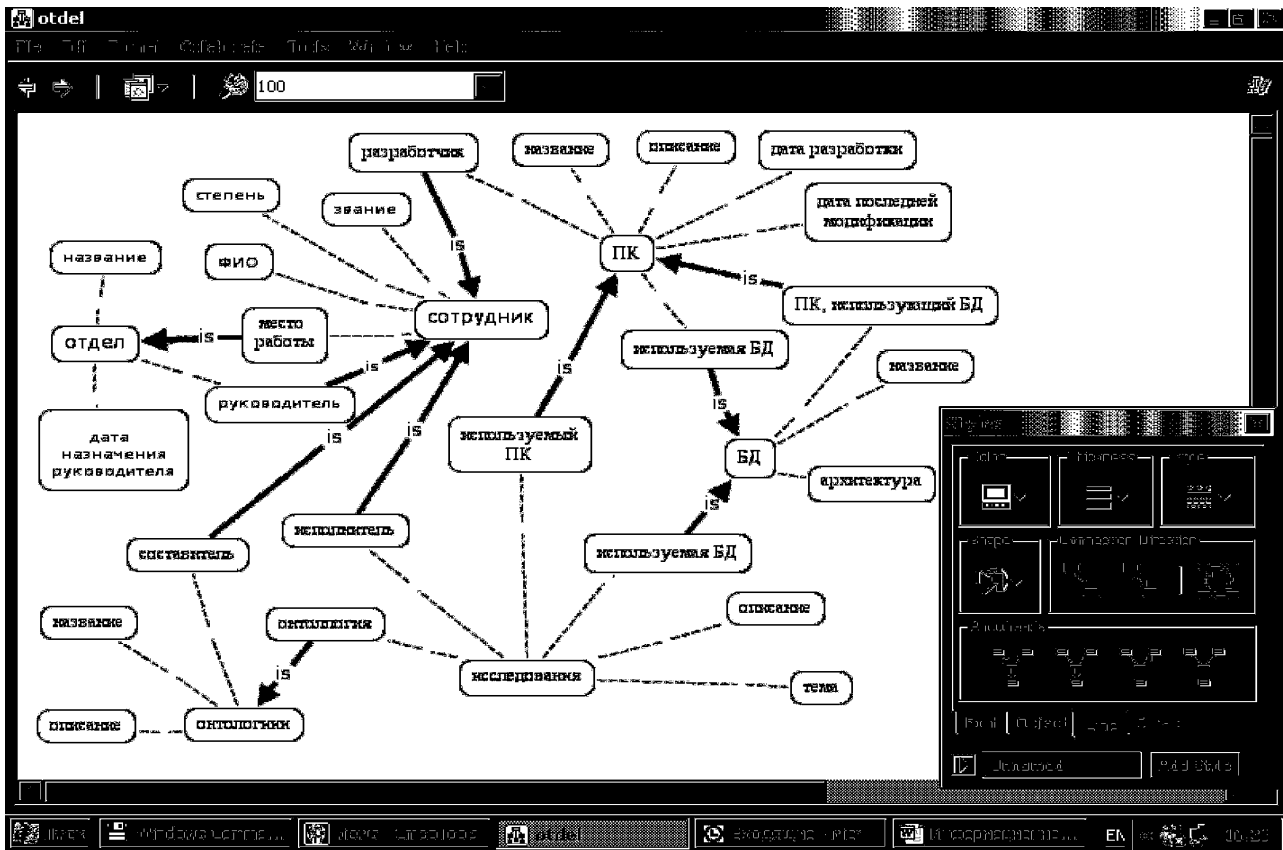


Рис. 4. Описание информационных ресурсов в системе SmartTools.

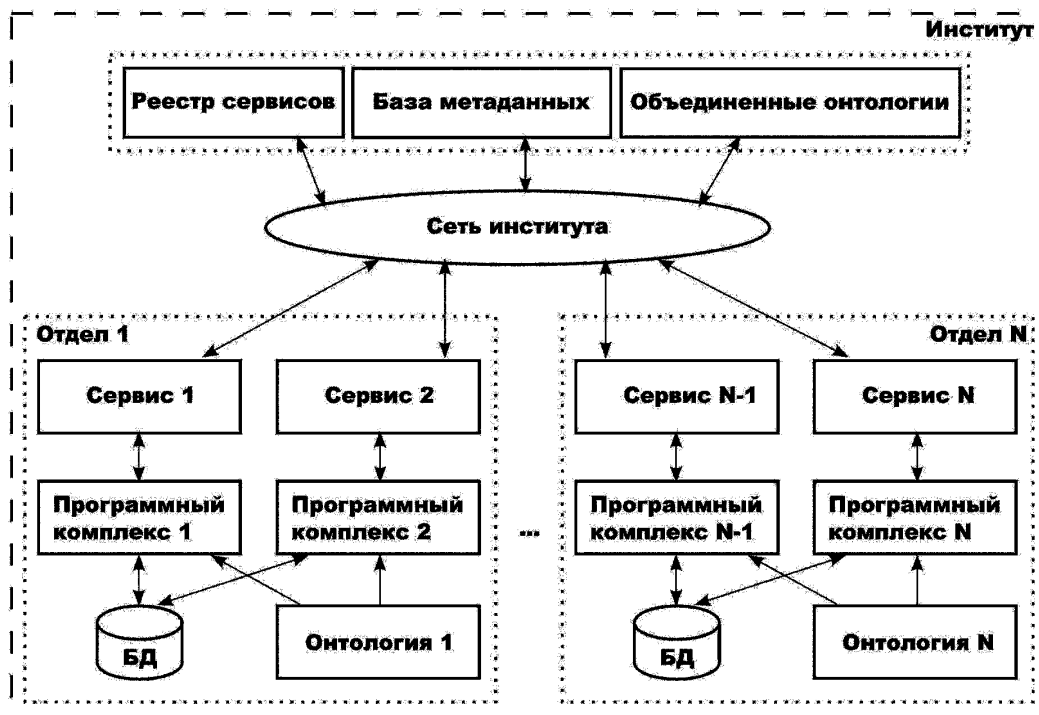


Рис. 5. Общая архитектура ИТ-инфраструктуры.

следующем для отображения метаданных информационных ресурсов института была использована инструментальная система StarTools [12], которая предлагает эффективный и интуитивно более понятный графический (или компонентный) способ разработки онтологий, когда пользователю предлагаются графические примитивы для отображения на экране концептов и связей между ними (рис. 4).

Общая архитектура ИТ-инфраструктуры приведена на рис. 5.

Заключение

Следует заметить, что предлагаемый авторами подход базируется на обобщении результатов, полученных при создании ИТ-инфраструктуры системных исследований в энергетике. Последние имеют свою специфику: основными инструментами исследований являются математическое моделирование и вычислительный эксперимент; исследования, проводимые в институте, тесно связаны и имеют общую информационную базу; за сорокапятилетнюю историю института создано большое количество программных продуктов и баз данных; более двадцати программных комплексов могут быть отнесены к промышленным и коммерческим продуктам и применяются не только в институте, но и в других научных и энергетических организациях России и за рубежом. Предлагаемый подход и полученные результаты могут быть полезны для любой научной организации, осуществляющей стратегию последовательного применения современных информационных технологий.

Список литературы

- [1] МАССЕЛЬ Л.В., БОЛДЫРЕВ Е.А., МАКАГОНОВА Н.Н. и др. Интеграция информационных технологий в системных исследованиях энергетики / Под ред. Н.И. Воропая. Новосибирск: Наука, 2003. 320 с.
- [2] МАССЕЛЬ Л.В. Методы и технологии создания ИТ-инфраструктуры научных исследований // Информационные и математические технологии в науке, технике и образовании: Тр. X Байк. Всерос. конф. Ч. 1. Иркутск: ИСЭМ СО РАН, 2005. С. 57–64.
- [3] КРИВОРУЦКИЙ Л.Д., МАССЕЛЬ Л.В. Фрактальный подход к построению информационных технологий // Информационная технология исследований развития энергетики. Новосибирск: Наука, 1995. С. 40–67.
- [4] БОЛДЫРЕВ Е.А. Вычислительная инфраструктура научных исследований и подход к ее реализации // Информационные и математические технологии в науке, технике и образовании: Тр. X Байк. Всерос. конф. Ч. 1. Иркутск: ИСЭМ СО РАН, 2005. С. 65–72.
- [5] КОПАЙГОРОДСКИЙ А.Н., МАКАГОНОВА Н.Н., ТРИПУТИНА В.В. Информационная инфраструктура научных исследований // Там же. С. 72–80.
- [6] FOSTER I., KESSELMAN C., TUECKE S. The anatomy of the grid: enabling scalable virtual organizations // Intern. J. Supercomputer Appl. 2001. Vol. 15, N 3.
- [7] FOSTER I., KESSELMAN C., NICK J., TUECKE S. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration // Open Grid Service Infrastructure WG, Global Grid Forum, June 22, 2002.

- [8] CHANNABASAVAIAN K., HOLLEY K., TUGGLE E.M. Migrating to a service-oriented architecture. IBM, Dec. 2003.
- [9] ФЕЙГИН Д. Концепция SOA // Открытые системы. 2004. № 6. С. 14–18.
- [10] ДУБОВА Н. SOA: подходы к реализации // Там же. С. 19–25.
- [11] ГАВРИЛОВА Т.А., ХОРОШЕВСКИЙ В.Ф. Базы знаний интеллектуальных систем. СПб.: Питер, 2000. 384 с.
- [12] IHMC SmartTools. <http://smart.ihmc.us>.

Поступила в редакцию 29 октября 2006 г.