

ИНТЕГРАЦИЯ РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННЫХ И ИНТЕЛЛЕКТУАЛЬНЫХ РЕСУРСОВ ДЛЯ ИССЛЕДОВАНИЙ ЭНЕРГЕТИКИ *

Л.В. МАССЕЛЬ

Институт систем энергетики им. Л.А. Мелентьева СО РАН

Иркутск, Россия

e-mail: massel@isem.sei.irk.ru

А.Н. КОПАЙГОРОДСКИЙ

e-mail: digger@istu.edu

В статье рассматриваются современное состояние и направления работ в области интеграции распределенных информационных и интеллектуальных ресурсов для комплексных исследований энергетики, проводимых в Институте систем энергетики им. Л.А. Мелентьева СО РАН. Необходимость такой интеграции обусловлена общностью и взаимосвязанностью исследований отдельных систем энергетики и топливно-энергетического комплекса в целом, общей информационной базой этих исследований. Первым этапом интеграции явилось создание ИТ-инфраструктуры системных исследований в энергетике, которая была разработана сотрудниками лаборатории информационных технологий в энергетике ИСЭМ СО РАН. Следующим этапом является создание онтологического пространства знаний и его интеграция в ИТ-инфраструктуру. В статье описываются предлагаемый методический подход, базовые технологии и разработанные инструментальные средства для реализации предложенного подхода. Работа выполнена при частичной финансовой поддержке грантов РФФИ №10-07-00264, №11-07-00192 и гранта программы Президиума РАН №2.29.

Введение. Институт систем энергетики им. Л.А. Мелентьева (ИСЭМ) СО РАН является признанным лидером в области системных исследований в энергетике. Основными научными направлениями ИСЭМ СО РАН являются: 1) теория создания энергетических систем, комплексов и установок и управления ими; 2) научные основы и механизмы реализации энергетической политики России и ее регионов. В рамках этих направлений выполняется комплексное рассмотрение исследуемых проблем и интегрируются исследования: систем энергетики (электроэнергетических, тепло-, газо-, нефте-, нефтепродуктоснабжения); энергетической безопасности России; региональных проблем энергетики; взаимосвязей энергетики и экономики; перспективных энергетических источников и систем; прикладной математики и информатики [1]. Объектами исследований в энергетике являются отраслевые (функциональные) энергетические системы: электро-, газо-, нефте- и угле- и теплоснабжения, а также объединяющая их общенергетическая система – топливно-энергетический комплекс (ТЭК) страны. В последнее время в России специфическими объектами исследований стали рынки топлива и энергии

* Работа выполнена при частичной финансовой поддержке грантов РФФИ №10-07-00264, №11-07-00192 и гранта программы Президиума РАН №2.29

разного территориального уровня. Особое место в комплексных исследованиях энергетики занимает решение сложных комплексных межотраслевых и междисциплинарных проблем, в том числе глобальных. С точки зрения управления развитием и функционированием системы энергетики (СЭ) являются сложными физико-техническими и территориально-распределенными системами (протяженность электроэнергетических, нефте- и газоснабжающих систем достигает десятков и сотен тысяч километров). Помимо территориальной распределенности и сложности, СЭ, входящим в ТЭК, присущ ряд общих свойств, среди которых выделяют: непрерывность и инерционность развития; непрерывность функционирования и взаимосвязь режимов работы элементов системы; многоцелевой характер и практическую невозможность полного отказа системы; неравномерность процессов потребления продукции; подверженность крупномасштабным внешним возмущениям (преднамеренным и непреднамеренным; возможность каскадного развития аварий; зависимость пропускных способностей связей (линий передач, трубопроводов) от их местонахождения, режимов работы и состава работающего оборудования; активное участие человека в процессе управления; неполноту (недостаточную достоверность) информации о параметрах и режимах работы системы; иерархичность. Перечисленные факторы позволяют отнести задачи, возникающие в исследованиях энергетики, к классу больших задач. Основными инструментами исследований энергетики являются математическое моделирование и вычислительный эксперимент. Исследования, проводимые в институте, тесно связаны и имеют общую информационную базу; для исследований ТЭК требуются агрегированные данные, которые, как правило, являются обобщением результатов исследований отраслевых систем; результаты исследований направлений развития ТЭК должны учитываться при исследованиях развития отраслевых систем энергетики. Таким образом, для проведения системных исследований энергетики необходимы координация и согласование исходной и результирующей информации для получения обоснованных выводов и рекомендаций, которые готовятся для внешних организаций, а для этого необходимо создание интегрированной информационной и вычислительной среды исследований, которой явилась разработанная в ИСЭМ СО РАН ИТ-инфраструктура научных исследований.

ИТ-инфраструктура системных исследований в энергетике. Под ИТ-инфраструктурой понимается телекоммуникационная распределенная информационно-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 разработки информационных и программных продуктов в области исследований энергетики, обмена ими и их представления на информационный рынок [2]. ИТ-инфраструктура, с одной стороны, является интеграционной информационной и вычислительной средой для проведения исследований в энергетике. С другой стороны, ее разработка создает предпосылки для поэтапного перехода: к созданию распределенных баз данных и программных комплексов; распределению и параллеливанию вычислений; созданию Web-ориентированных программных комплексов (вычислительных серверов); оказанию информационных услуг на основе наукоемких информационных и программных продуктов (создание Web-служб и Web-сервисов) [2, 3, 4, 5]. ИТ-инфраструктура включает интеграционную информационную инфраструктуру, распределенную вычислительную инфраструктуру и телекоммуникационную инфраструктуру (последняя в этой статье не рассматривается). В настоящее время ведется работы по развитию ИТ-инфраструктуры, а именно дополнение ее интел-

лектуальными ресурсами, которые, по аналогии с информационной и вычислительной инфраструктурами, условно объединены в интеллектуальную инфраструктуру. Общая схема ИТ-инфраструктуры исследований в энергетике приведена на рис. 1.

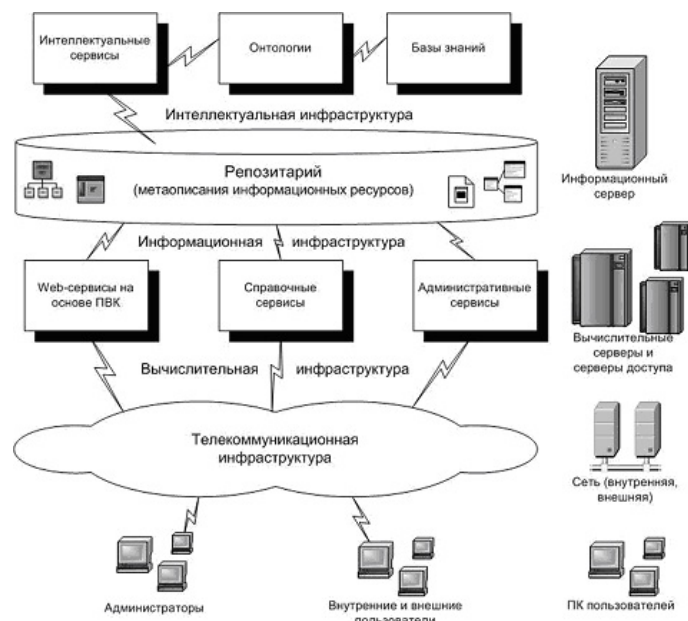


Рис. 1. ИТ-инфраструктура исследований в энергетике

Информационно-вычислительные и интеллектуальные ресурсы.

Информационные ресурсы объединяют:

- данные о сотрудниках, проводимых ими научных исследованиях и их результатах. информацию обо всех разрозненных базах данных и хранилищах
- данных, используемых для хранения базовой (основной), промежуточной и результирующей информации, необходимой для исследований,
- информацию о методах (алгоритмах), выполняющих какие-либо действия над информацией (программных комплексах, пакетах прикладных программ, различных сервисах).

Под *интеллектуальными ресурсами* понимаются знания, которыми обладает институт, представленные как в традиционном виде (электронные тексты статей, отчетов и монографий), так и в виде моделей данных и моделей программ и онтологий предметных областей энергетики [6]. Модели данных представляются либо в виде инфологических моделей (ER-моделей), в нотации Чена (подготовленные, например, в Microsoft Visio), либо в виде моделей данных, подготовленных с помощью CASE-средств, например, ErWin. Модели программ могут быть описаны либо на языке UML, либо с помощью онтологий, представленных на языках OWL или XML. *Онтологии* – одно из современных направлений в области искусственного интеллекта. В общем виде онтологии определяют как базу знаний специального вида, или как «спецификацию концептуализации» предметной области. Это значит, что в предметной области на основе классификации базовых терминов выделяются основные понятия (концепты) и устанавливаются связи между ними. Этот процесс называют концептуализацией. Затем онтология может быть представлена в графическом виде или описана на одном из формальных языков (формальная онтология) – это процесс спецификации онтологий. В основе представления

данных в ИТ-инфраструктуре лежат онтологии.

Информационная инфраструктура интегрирует как информационно-вычислительные, так и интеллектуальные ресурсы. В информационной инфраструктуре выделяются три уровня (слоя): уровень модели метаданных информационной инфраструктуры, уровень метаданных и данных ИТ-инфраструктуры. Для построения модели информационной инфраструктуры используются онтологии, объекты ИТ-инфраструктуры описываются метаданными на основе созданной онтологии [7, 8, 9]. Реализация всех программных компонентов выполняется в рамках SOA, центральным информационным компонентом является Репозиторий, реализация которого выполнена на базе СУБД Firebird и Microsoft SQL Server, функциональность СУБД расширяется программным ядром. Кроме Репозитория также реализованы программа администрирования и специализированный драйвер, программа извлечения данных, программа извлечения файлов, приложение для отображения информации в Internet и др. Для выполнения административных задач была разработана программа администрирования, которая позволяет достаточно легко проводить первоначальное проектирование и изменение структуры метаданных, вносить и модифицировать описания информационных объектов Репозитория, а также дает возможность извлекать данные из Репозитория в различные форматы. Для выполнения запросов к базам данных создано специальное приложение, которое базируется на драйвере Репозитория. Основной целью при его разработке было построение такого приложения, которое бы давало возможность пользователям извлекать данные для своей работы из различных баз данных в удобном формате. Извлечение данных можно осуществляться из целого ряда СУБД: MySQL, PostgreSQL, Firebird, InterBase, Microsoft SQL Server, Oracle, Sybase ASE, через драйвера OLE-DB, ADO и ODBC. Результаты выполнения запросов представляются пользователю в виде таблиц и могут быть сохранены в виде текстовых файлов, файлов Microsoft Excel, dbf-файлов или в виде других баз данных. Возможно построение на основе полученных данных OLAP-кубов, для этого пользователю необходимо указать изменения куба (dimension), меры (measure) и ограничения данных. В Репозитории запросы хранятся в виде отдельных объектов с указанием ссылок на базы данных, в качестве атрибутов этих объектов хранятся названия запросов, SQL-скрипты и списки параметров. Также поддерживается построение пользовательских (произвольных) SQL-запросов. Технологию использования информационной инфраструктуры можно разбить на четыре этапа: *построение модели метаданных* – на этом этапе определяются «правила хранения» метаданных, *внесение метаданных* – производится описание информационных ресурсов в информационной инфраструктуре, *извлечение метаданных и извлечение данных* – использование инфраструктуры для поддержки проведения исследований как интегрированного источника получения информации [9].

Хранилище данных и знаний исследований систем энергетики. Для поддержки исследований отдельных систем энергетики авторами предлагается использовать специализированные хранилища данных и хранилища знаний для каждой системы энергетики [10]. Под знаниями о предметной области в первую очередь понимаются декларативные явные знания [11], но в системе также предусмотрена возможность хранения процедурных знаний (описания программ и алгоритмов). Метаданные позволяют описывать знания, выполнять их классификацию и каталогизацию, и используются для быстрого и удобного поиска. При применении «типовых решений» поддержки исследований отдельных систем задача построения единого корпоративного хранилища для решения комплексных проблем энергетики значительно упрощается.

В хранилище данных и хранилище знаний отдельной отрасли энергетики словарь предметной области является общим (одним) и содержит свойственные ей классификаторы (рис. 2). Метаданные также являются общими и описывают как структуру данных, так и документы, помещаемые в хранилище. Таким образом, хранилище данных и знаний состоит из четырех основных логических частей: словаря предметной области, метаданных, непосредственно данных хранилища, которые физически расположены в базе данных, и декларативных знаний, представленных в виде документов, которые находятся в файловом хранилище. Стоит отметить, что ограничения накладываются только на структуру метаданных, которые описывают документы, находящиеся в хранилище, модели словаря предметной области и хранилища данных. Процесс исследования любой энергетической системы начинается со сбора массива исходных данных, который может быть получен из различных статей, отчетов, статистических сборников, также в качестве исходных данных могут выступать результаты предыдущих исследований. Внесение информации выполняется с привязкой к словарю предметной области: исследователь должен выполнить сопоставление определенных отчетных или статистических данных с регионом, категорией ресурса, его целевым назначением, должен указать и другие классификационные характеристики.

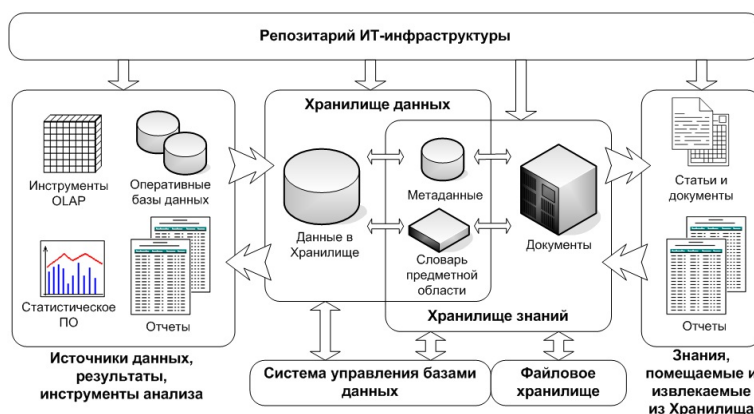


Рис. 2. Архитектура хранилища данных и знаний исследований энергетики.

В Репозитории ИТ-инфраструктуры описываются хранилища данных и знаний отдельных систем энергетики, указывается их расположение (адреса серверов) и интерфейсы взаимодействия, описаны и оперативные базы данных, используемые в исследованиях, программные комплексы, научные труды сотрудников института и др.

Для конфигурирования хранилища данных и описания модели исследуемой системы энергетики была разработана программа администрирования. Просмотр, корректировка и извлечения данных из хранилища выполняется с помощью клиентской программы, которая в своей работе активно использует метаданные, расположенные в хранилище. Для работы с хранилищем знаний была реализована программа, которая позволяет вносить, описывать, находить и извлекать декларативные знания, представленные в виде документов. Реализация программного обеспечения выполнена на объектно-ориентированном языке Java (Java Standard Edition) в среде NetBeans. В настоящее время в качестве базовой СУБД используется Firebird. Стоит отметить, что при выбранном подходе к созданию инструментальных средств и использованию метаданных возможен достаточно легкий переход почти на любую другую СУБД.

Заключение. В статье описан методический подход к интеграции информационных и

интеллектуальных ресурсов системных исследований энергетики. Необходимость интеграции обусловлена взаимосвязанностью и единой информационной базой исследований, а также необходимостью сохранения интеллектуальных ресурсов института. Интеграция осуществляется в рамках информационной инфраструктуры, являющейся важным компонентом созданной сотрудниками лаборатории информационных технологий в энергетике ИСЭМ СО РАН ИТ-инфраструктуры научных исследований. Предложенный и реализованный методический подход основан на использовании онтологий – современной формы представления знаний, причем онтологии предлагается использовать как для представления энергетических знаний и их описания в информационной инфраструктуре, так и для проектирования современных программных комплексов. Исследования, описанные в статье, выполнены при частичной финансовой поддержке грантов РФФИ №10-07-00264, №11-07-00192 и гранта программы Президиума РАН №2.29

Список литературы

- [1] Беляев Л.С., Санеев Б.Г., Филиппов С.П. и др./ под ред. Воропая С.П. Системные исследования проблем энергетики. Новосибирск: Наука, 2000. 558 с.
- [2] Массель Л.В. ИТ-инфраструктура научных исследований и открытая образовательная среда // Вестник ИрГТУ, 2005. – №4 – С. 9-15.
- [3] Массель Л.В., Болдырев Е.А., Макагонова Н.Н., Копайгородский А.Н., Черноусов А.В. ИТ-инфраструктура научных исследований: методический подход и реализация // Вычислительные технологии, т.11, 2006.- С.59-67.
- [4] Воропай Н.И., Массель Л.В. ИТ-инфраструктура системных исследований в энергетике и предоставление ИТ-услуг. – Известия АН – Энергетика, №3, 2006.- С. 86-93.
- [5] Massel L.V., Kopyagorodsky A.N., Chernousov A.V. IT-infrastructure of research activities realized for the power engineering system researches // X International Conference “Computer Science and Information Technologies”, 2008.- Turkey, Antalia.- P. 106-111.
- [6] Массель Л.В., Копайгородский А.Н. ИТ-инфраструктура научных исследований как основа информационного обеспечения научно-образовательного процесса / Труды VII Всероссийской научно-практической конференции «Корпоративные библиотечные системы: технологии и инновации», 2009. - Спб: СпбГПУ. – С. 88-93
- [7] Массель Л.В., Копайгородский А.Н. Разработка и интеграция основных компонентов информационной инфраструктуры научных исследований // Вестник ИрГТУ.-2006.- № 2 (26), т.3.- С.23-29.
- [8] Массель Л.В., Копайгородский А.Н. Технологии и система хранения данных и знаний для исследований в энергетике // Материалы Всероссийской конференции «Современные информационные технологии для научных исследований».- Магадан: СВНЦ ДВО РАН, 2008.- С. 64-66.
- [9] Копайгородский А.Н. Методы, модели и программные средства построения информационной инфраструктуры исследований в энергетике / Автореферат дисс. на соискание степени канд. техн. наук.- Иркутск: ИСЭМ СО РАН, 2008.- 25 с.
- [10] 10. Массель Л.В., Осама Ель Сайед Шета, Копайгородский А.Н. Разработка хранилища данных и знаний для поддержки исследований энергетики // Вестник ИрГТУ. - 2010. – № 5 (45). – С. 11-16
- [11] Тузовский А.Ф., Чириков С.В., Ямпольский В.З. Системы управления знаниями (методы и технологии) / под ред. В.З. Ямпольского. – Томск: Изд-во НТЛ, 2005. – 260 с.